Patient Satisfaction in General Medical Practice And The Association With Patient Shortage

Christina Sivertsen

Master Thesis

as part of the Master of Philosophy Degree in Health Economics, Policy and Management

UNIVERSITY OF OSLO

Department of Health Management and Health Economics

May15th, 2014

2014

Patient Satisfaction in General Medical Practice And The Association With Patient Shortage

Christina Sivertsen

http://www.duo.uio.no

Print: Reprosentralen, Universitetet i Oslo

Abstract

Background**:** After the introduction of the general practitioner (GP) scheme in 2001, many GPs did not achieve their preferred list size, and experienced patient shortage. Patients have imperfect information about GPs. Results in the literature have indicated that GPs with inferior quality are more likely to have patient shortage as patients choose their GP based on a perception of quality. Previous research has established that patient shortage is associated with lower patient satisfaction except satisfaction with waiting times. In 2012, the anonymous GP rating site Legelisten was established. This service enables people to rate the GP along various dimensions like listening skills, opening hours and waiting time.

Objective**:** To examine whether data from an anonymous rating site where the sample is based on self-selection yields similar results as the literature based on surveys of a randomized sample. More specifically, we investigate whether patient shortage is associated with lower patient satisfaction with the interpersonal relationship between patient and GP, and higher satisfaction with waiting times.

Method: The study employs descriptive statistics and multilevel ordinal regression analysis in order to investigate the association between patient shortage and patient satisfaction when accounting for influence by age, gender and competition.

Results**:** Patient shortage is associated with lower odds of high satisfaction with the interpersonal relationship, and higher odds of high satisfaction with waiting times compared with GPs who have full lists.

Conclusion**:** The results in the analysis based on a self-selected sample are aligned with the existing literature indicating that there is an association between patient shortage and lower patient satisfaction on dimensions describing the interpersonal relationship between the patient and the GP. GPs with patient shortage react by offering shorter waiting times, as satisfaction is more likely to be higher on those dimensions for the same GPs. The results suggest that the rating site could be useful in alleviating the information problem patients face when choosing a GP.

Acknowledgement

I would like to thank my supervisor Geir Godager for providing me with guidance and constructive and helpful feedback through this entire process of writing my master thesis.

I am grateful for spending two years with my fellow students, class of 2012. A special thanks goes to my study friends Stine Bergliot Olsen and Tone Olsen Moberg for our daily lunch breaks with filled with frustration, discussions and laughter at all times.

I would also like to thank my close friends, especially my dear Øystein and the Tuesday-ladies. They have supported me with proofreading, encouragement and showing me a good time when a break was needed.

Christina Sivertsen
Oslo, May 2014

Abbreviations

GP	General Practitioner
MLE	Maximum Likelihood Estimation
OLS	Ordinary Least Square
NAV	Norwegian Labour and Welfare Administration
SSB	Statistics Norway
NSD	Norwegian Social Science Data Service
NO	Number Of
CI	Confidence Intervals
SE	Standard Error
HHI	Herfindahl-Hirchman Index
Log HHI	Log of Herfindahl-Hirscman Index
NHS	National Health Service (UK)
VIF	Variance Inflation Factor

Table of Contents

List of Figures

List of Tables

1 Introduction

After the introduction of the regular general practitioner (GP) scheme in 2001, many GPs did not achieve their preferred patient list size and experienced patient shortage. Economic theory has described how these GPs behave in the market in order to attract additional patients by providing income-generating services and shorten their waiting times (Godager & Lurås, 2009; Iversen & Lurås, 2000, 2002; Scott, 2000). Patients choose GPs with imperfect information, but may be influenced by their perception of GP quality. The patients perception of GP quality might influence which level of demand the GP faces; patient shortage or filled capacity (Biørn & Godager, 2010; Iversen & Lurås, 2011). Literature shows that patient satisfaction with their GP tend to be high (Williams, Coyle, & Healy, 1998). Studies in the Norwegian context have found that patient satisfaction is lower with the interpersonal relationship, and higher with waiting times, when the GP face patient shortage (Godager & Iversen, 2010, 2014; Lurås, 2007).

The aim of this study is to explore whether the data applied from an anonymous rating website can confirm and contribute to already existing results generated from the Norwegian survey of living condition (Godager & Iversen, 2010, 2014; Lurås, 2007). More specifically, the objective is to analyze the association between patient shortage and patient satisfaction on dimensions of the GPs practice describing the interpersonal relationship and time use.

The thesis is structured as follows: Chapter one describes the background of the term patient shortage in the Norwegian context followed by theory regarding the market, quality and expected GP behavior when facing patient shortage. Following, how patient satisfaction can measure quality and its association with patient shortage is presented before stating the working hypotheses. Chapter two presents the applied data and relevant variables for the thesis followed by a description of the analysis method and the empiric model. Chapter three presents descriptive statistics and results from the regression analysis. Chapter four provides a discussion of the results in light of existing research findings, further research and implications, and will be summarized with a conclusion.

1.1 Background

In 2001, the general practitioner reform was implemented with the leading purpose to improve the quality in primary health care (Ot.prp.nr.99). With this reform patients were given a right to be listed with a regular GP, and 77% became enlisted with their first choice (Finnvold, Paulsen, & Lurås, 2003). The allocation process was a result of the patients preferred choice and the GPs preference of patient list size- which they report to the Norwegian Labour and Welfare Administration (NAV). When the allocation process was complete, many GPs got fewer patients listed than they preferred. The reason why some GPs experienced patient shortage, was that not all of them were a popular first choice listing, and in addition there was an 18% increase in physician man-years in order to ensure sufficient availability (Grytten, 2009). The GP payment system consists of 30% per capita, and 70% refund and patient copayment (Ot.prp.nr.99). Le Grand (2007) argues that such adjustments create a situation where the GPs can compete for patients. When we have a competitive market combined with a free choice of GP, certain conditions have to be met to ensure that incentives for quality of care are present. There should be excess capacity, a sufficient amount of GPs with open lists to ensure that patients have a real choice, and available information to base the choice upon. The physicians should be provided with incentives to compete for a higher demand through quality, such as per capita payment. Further, the patients must be responsive to variation in quality, and choose their GP based on this information (Le Grand, 2007).

"*More specifically, for choice to act as an effective driver of quality, it is necessary to rely upon the users judgment about the quality and responsiveness of the service and for providers to react to choices made on the basis of those judgments"* (Le Grand, 2007, p. 117).

The Norwegian Directorate of Health has documented that people have fewer available choices in the GP market now, compared to the beginning of the reform when there was a high proportion of GPs experiencing patient shortage (Gaardsrud, 2012). The percentage of GPs with open lists has decreased from 2001 to 2012 from 59.6% to 40. 1% respectively. In the same period, the proportion of GPs with patient shortage has decreased from 11.2% to 5.5%. Ordinary GP switching, which mean a switch of GP without changing the address, has also decreased from 5.3% to 1.6% and is highest in Oslo with 2.2%. In the rural areas on the other hand, the proportion of municipalities with fewer than two GPs with open lists has decreased from 146 to 116 (Gaardsrud, 2012). Regarding the available information about quality, there has not been any substantial change since the reform. It is a common belief that patient

choice might drive quality. However, there exist few indicators for patients to base their choice upon. Although the government has been monitoring certain quality indicators and patient satisfaction with the GP scheme, none of this information is public. Patients may choose a GP based on information of list size and the number of vacancies on the list.

After the private rating site Legelisten.no was launched in 2012, many people have chosen to rate their GPs. The ratings provide people with information relevant for choosing a GP dependent on whether they trust the information provided. Further it is an additional information source regarding the association between patient satisfaction and patient shortage, which have been explored earlier based on the Norwegian survey of living conditions (Godager & Iversen, 2010, 2014; Lurås, 2007). Although the previous studies are of a different character, satisfaction measures on Legelisten are aligned with, and contribute to these results.

1.2 Demand and quality in the market for general practitioners

The market for health care services deviates from a perfectly competitive market. Pauly and Satterthwaite (1979) refer to services provided by GPs as reputation goods. A market consists of reputation goods if it consists of differentiated services, and consumers gain information of the services through its reputation in their network. The primary market is normally thought to be monopolistic competitive as the providers can set the price and influence quality attributes within a limited range, and consumers have imperfect information about the providers and their services. When the number of providers increase, consumers have less information on the individual GPs (Pauly & Satterthwaite, 1979). McCarthy (1985) hypothesizes that competition alleviates the information problem because it drives the providers to increase the amount of information, and reports significant findings where patients are sensitive to price, scheduling practice and quality. If patients are able to distinguish between quality, then demand and price is likely to be higher for GPs who are of higher quality (Dranove & White, 1986).

The effects of competition on quality can take two directions. Gaynor (2006) argues that when the market is characterized by price regulation, providers must compete through quality. It is further pointed out that this combination of competition and price regulation can result in high quality (Gaynor, 2006). This theory is supported by several studies focusing on the market of hospitals (Gaynor, Moreno-Serra, &

Propper, 2013). Dranove and Satterthwaite (2000) argue that when patients search for a physician of their choice, the information about quality may be noisy. Patients are able to correct for contaminating noise in their observations, although not perfectly so. The more noise in an attribute, the less responsive the patient becomes.

Competition might not necessarily have a positive effect on quality. Patients have imperfect information of the providers, and have previously been associated with being poor judges of technical quality (Arrow, 1963; Dranove & White, 1986). Pauly (2004) discusses how the impact of competition on quality depends on the ability to judge quality and what incentives are given through the payment system. Propper, Burgess, and Green (2004) found that competition decreased quality after a reform in England, supporting the theory that when quality signals are weak or noisy, quality may be reduced. Dranove and Satterthwaite (1992) and Gaynor (2006) argue that competition may result in decreasing quality when the demand side is unresponsive to changes in quality. However, the long term relationship between a patient and a physician provides the patient with monitoring opportunities, which increases their information and ability to make judgments on both technical quality, waiting times and preferable characteristics. Because of this, it is in the physicians best interest to provide the best possible service in order to keep patients, and build up a positive reputation to increase demand (Dranove & White, 1986). Patients may accumulate knowledge about the physician with time and experience, reducing the information gap (Godager, Iversen, & Ma, 2012; Gravelle & Masiero, 2000; Scott, 2000).

Theory suggests that whether patients are able to judge quality and whether it affects demand are important empirical questions. Gravelle, Propper, and Santos (2013) did a study in England on whether an increase in quality affects demand among primary physicians. A portion of the physicians in one market area increased the quality on chosen measurable quality indicators. As a result, demand for those physicians increased by 15%. Amongst published and unpublished indicators, the patients were most responsive towards published quality indicators. Biørn and Godager (2010) have also conducted a study on whether quality affects patients demand. They assume that when information on GPs is limited, patients are not affected by e.g. mortality rates directly. But they hypothesize in their study that patients have perceptions of the latent variable quality, which influences their choice of GP. Their findings indicate that patients are responsive to a GPs quality level, and this

influences demand. Theory on how patients may be able to sort out noise and distinguish between different quality levels is illuminated below.

We may consider a market where N patients have imperfect information when choosing between two, with the exception of quality, identical GPs. One of the GPs supply services of high quality (H), and the other low quality (L), but due to imperfect information, it is not possible for the consumer to perfectly observe quality. Each GP has the capacity to enlist a maximum number of patients denoted by n^{max}. We assume that the two physicians are capable of serving the whole market, implying that $2n^{max}>N$. The demand side, consisting of N patients, has to make a choice between the two of them without having perfect information about which physician is of quality H or L. Noise and error cause the GP of low quality to achieve some demand for their service $0 < n_L$. But as experience and reputation relieve some of the imperfect information, the high quality GP will have higher demand $n_L<n_H$.

Biørn and Godager (2010, p. 843) describe quality as being distinguishable between true, unknown quality μ_j of GP J $(J = L, H)$ and perceived quality q_{ij} by the consumer i $(i = 1, .., N)$. The relation between the true and perceived quality may be expressed as $q_{ij}=\mu_j+\epsilon_{ij}$ where ϵ_{ij} denotes noise and measurement error. Conditional on μ_j the probability that consumer i chooses GP J can now be expressed as

$$P_{iH}=P(\mu_H+\epsilon_{iH}>\mu_L+\epsilon_{iL})=P(\epsilon_{iL}<\mu_H-\mu_L+\epsilon_{IH}) \quad (1)$$

The assumptions we make for the noise term ϵ_{ij} leads to different choice models. If we assume ϵ_{ij} to be independent and to follow an extreme value distribution of type 1, we will consider choice probabilities consistent with the logit type model. Following, we get the expression, similar to Train (2003, p. 78);

$$P_{iH}=\frac{e^{\mu H}}{e^{\mu H}+e^{\mu L}} \quad P_{iL}=\frac{e^{\mu L}}{e^{\mu H}+e^{\mu L}} \quad (2)$$

From this model we may express expected demand for a GP of high quality as NP_{iH} with the following predictions: GPs of higher quality have a higher probability of being chosen by consumers, $P_{IH}>P_{IL}$ and therefore also higher likelihood of achieving the maximum list size $n_H=n^{max}$ than what is the case for the GPs of low quality. Then it logically follows that GPs of low quality have a higher probability of facing patient shortage, $n_L<n^{max}$ (Biørn & Godager, 2010, p. 844). This situation might look as described by Becker (1991) where some firms may experience excess demand, which can be a sign of quality, and thus accumulate even higher demand. Firms like this

profit on keeping their prices fixed, and establishing a queuing system. The other firms of lower quality, experiencing patient shortage, profit on lowering their prices, creating multiple equilibria in the market.

1.3 GPs decision problem when facing patient shortage

In Norway, the physician market is regulated by fixed prices and many patients are exempt from co-payment. As a result, price cannot be used as a competitive device, and thus GPs must compete on quality (Gravelle & Masiero, 2000). The GP scheme encourages competition over patients in general, and especially GPs who experience patient shortage may seek to attract patients by improving accessibility, offering more services and shorter waiting times (Iversen & Lurås, 2002). The study of this decision problem is unique to the Norwegian context, as both the GPs preferred list size and actual patient list size is registered. This data allows us to describe GPs behavior with and without patient shortage.

1.3.1 Income motivated behavior

Economic models describing GP behavior come in many varieties. Common for many models is that they include economical incentives, and preference for leisure as part of the physician objective function. They often include an ethical argument as well (Scott, 2000). Iversen and Lurås (2000), Iversen (2004), Iversen and Lurås (2002) and Godager and Lurås (2009) have contributed to the literature on income motivated behavior relevant for the Norwegian market, when GPs are facing patient shortage. Income motivated behavior is a term for when a physician, facing patient shortage, provides more service compared to unconstrained colleagues, and when this high intensive service is a result of the deficit in their preferred practice profile. GPs preferred list size reveals their preferred practice style. A preferred list size higher than the actual list size indicates a wish of less leisure and a higher workload, possibly leading to less time with each patient. When a GP faces patient shortage, she has more motivation for adjusting the variables under her control. The GP may provide more services, and devote more time to each patient (Iversen, 2004; Iversen & Lurås, 2000).

1.3.1.1 Service provision

The three papers (Godager & Lurås, 2009; Iversen, 2004; Iversen & Lurås, 2000) all include an ethical argument in their models. The ethical argument describes how the

appropriate service intensity is not always clear, but is constrained within the ethical conduct, thus we have a gray area in the acceptable range of service provision. Marginal health benefit of health care service is assumed to be a declining function of service volume, but within the gray area, each practice profile is considered equally appropriate. It is only within this gray area that income motivated behavior applies. In the models, the physicians practice profile can be described by the number n of patients on the list, and n^{max}, which is the maximum number that may be enlisted. The physician may serve less than the maximum number of patients $0< n \leq n^{max}$. If the physician wishes to enlist more patients, then the constraint becomes binding at $n = n^{max}$ and the physician faces patient shortage. The gray area is included in the models by limiting service volume to a certain bounded range $[S_1, S_2]$, where $S_1< S_2$ and are exogenous to the physician. It is assumed that the provided service volume s is produced until the marginal effect equals zero and must be within the interval $S_1 \leq s \leq S_2$. When we have a combined payment system consisting of a capitation system and a fee for service system, the size of capitation relative to fee for service decides the optimal service provision. For a physician not facing patient shortage, an additional service has an opportunity cost of reduced income that may have been collected from capitation. When this opportunity cost becomes significant, the optimal value of services will be $s=S_1$, where only a minimum of services will be provided. The physician is able to keep the payoff from ns constant by reducing s and raising n. Physicians facing patient shortage, have lower opportunity costs relative to capitation because of lower n and have incentives to increase service volume closer to $s=S_2$ (Godager & Lurås, 2009; Iversen, 2004; Iversen & Lurås, 2000). Iversen and Lurås (2000) conclude that physicians with patient shortage generated higher income and more frequent consultations and laboratory tests than unconstrained physicians. This hypothesis gains further support in Iversen (2004) where the long term effect after five years, showed that constrained physicians provided 15 % more services per patient. Godager and Lurås (2009) found that GPs who experience a shortage of patients provide more community health service compared to their colleagues. Kann, Biørn, and Lurås (2010) studied service and prescription intensity amongst elderly and found that GPs under competition and facing patient shortage provided more consultations and services compared to their colleagues.

1.3.1.2 Waiting time

The physicians optimal waiting time is the focus of several papers. Previous international studies have shown that it is more favorable for the physicians to have longer waiting times than short. This is modeled through the physicians costs, which increase with the number and length of the consultations, and decrease with longer waiting time. The waiting time tends to become shorter when there is more competition between physicians (Mueller, 1985; Sloan & Lorant, 1977). Godager and Iversen (2010) have provided evidence that support the theory that competition reduces waiting time. They argue in their analysis of the Norwegian living condition survey, that those counties with a high proportion of GPs with patient shortage, cause a competitive situation where the GPs have extra incentives towards reducing waiting times in order to avoid GP switching. However this effect changed over time. At the end of the period of 2002 to 2008, there was no longer a statistical difference in waiting times on municipality level dependent on the proportion of the GPs with patient shortage in each county. In the following survey in 2012, waiting time within municipalities with a higher share of GPs experiencing patient shortage was significantly lower (Godager & Iversen, 2014).

Iversen and Lurås (2002) have modified the model for optimal waiting time according to the Norwegian context. In their model, the GP has a choice problem concerning leisure and income. When the price is given, they may choose their preferred patients n, the intensity of service provision s and waiting time w. Service volume s is assumed to be dependent on waiting time w, i.e. when a practice has long waiting times, there will be less service per patient as patients seek treatment elsewhere or recover on their own. Waiting time is dependent on leisure and the amount of patients. The optimal level of patients and leisure is determined by the marginal utility for leisure, which depends on a mix of capitation and fee for service. Waiting time is assumed to be a function of the number of patients and leisure but is constrained by a maximum wait of $0 < w(n, l) \leq w^{\max}$. The optimal number of patients and leisure time is further constrained by the demand, which mainly consists of exogenous arguments such as GP characteristics combined with the composition of other GPs in the market. When the GPs funding is a mixed system, high n is associated with maximum waiting time $w^{\max}$. Low n is associated with low w. When a significantly sized shortage of patients emerges, the GP has an involuntary increase of leisure. Since the GP has preferences regarding income and

leisure, they have some choices. When the marginal utility of income is higher than the marginal utility of leisure, the optimal waiting time may be reduced in order to increase income from fees, and attract new patients to the list, thereby increasing income from capitation. In Iversen and Lurås (2002), the physicians, who do not face a patient constraint, offer a longer waiting time. This is driven by the stochastic demand for consultation; by increasing the waiting time, a smooth demand and higher remuneration per hour may be achieved. Iversen and Lurås (2002) found that physicians experiencing patient shortage, offered an almost 40% lower waiting time, and these physicians also experienced an increase in enlisted patients the next period. An interpretation of this observation is that their effort towards attracting new patients was successful. Iversen and Lurås (2002) conclude that unconstrained physicians are considered to provide higher quality services compared to physicians with patient shortage. Thus only physicians considered to be of inferior quality choose to compete by reducing waiting time. Therefore, as physicians vary in quality, in the eyes of the patients, we can expect a market situation with a wide range of waiting times.

1.4 Patient satisfaction as a proxy for quality

"Like beauty, quality is at least partly in the beholders eye. It has two aspects; quality of action and quality of perception" (Vuori, 1991, p. 189). How to best assess quality in healthcare has gained much attention. Traditionally, quality has been assessed addressing the structure, process and outcome from the experts' perspective. In the recent years, there has been a movement towards including the patient perspective (Chow, Mayer, Darzi, & Athanasiou, 2009; Sullivan, 2003). The three categories of, structure, process and outcome were first introduced as the source of inference about quality in healthcare by Donabedian (1966). Structure denotes the material, human resources and organizational structure in the relevant setting. Process denotes how and what diagnosis and treatment was implemented. Outcome denotes the health status after receiving treatment, and if broadly defined, includes patient satisfaction (Donabedian, 1966, 1988). Donabedian (1988) further explains that patient satisfaction may be considered to be part of the health status and a desired outcome of quality care. Whether a patient is satisfied or dissatisfied provides information in all the categories for quality assessment, technical quality, as well as the interpersonal relationship between the patient and physician. The interpersonal quality might be equally important as the technical quality, as it is the base for

successful diagnosis and implementation of treatment. Therefore, patient satisfaction should be considered a valid quality indicator no matter its strength and weaknesses (Donabedian, 1988). The view that patient satisfaction is a valid indicator for quality has gained a massive support over the years (Aharony & Strasser, 1993; Ford, Bach, & Fottler, 1997; Hekkert, Cihangir, Kleefstra, van den Berg, & Kool, 2009; Sitzia & Wood, 1997; B. Williams et al., 1998). Vuori (1991) argues, that patient satisfaction is not only a proxy for quality, but also an essential part of quality. However, some have claimed reasons for being skeptic. Patients ability to judge the quality of healthcare has been questioned, and this has implications for whether the satisfaction of health care is a measure of quality health care (Pascoe, 1983). Some of the concerns have been that satisfaction is merely a reflection of individual emotional need, rather than quality of care and does not correspond to an objective measure. The deviance from objective measures is not necessarily a weakness however. It is the core strength of satisfaction measures, as we seek to capture something more. Together with the choice of GP, empirical evidence points to patients being sensitive to the difference between provider qualities and mere courtesy, and make their satisfaction rating with this in mind. In fact, the patients' satisfaction tends to be aligned with the professionals' own evaluation of quality. Patient satisfaction contributes with a valuable perspective that balances the objective structure, process and outcome, and serves as an outcome measure of the quality of the received health care (Pascoe, 1983; Ware, Snyder, Wright, & Davies, 1983).

1.4.1 Components and determinants influence patient satisfaction

Patient satisfaction is a complex concept widely debated in the literature. Numerous different definitions have been proposed with different emphasis on what it actually consists of, and what it captures (Sitzia & Wood, 1997; Williams, 1994; Williams et al., 1998). Chow et al. (2009, p. 436) describe patient satisfaction in terms of healthcare simply as *"the degree to which a patient feels they have received high-quality of care. If a patient feels they have received high-quality care, they are more likely to be satisfied"*. With a simple definition such as this, it must be further pointed out that patient satisfaction is a subjective and a twofold measure consisting of satisfaction components and satisfaction determinants (Ware et al., 1983). Satisfaction components denote the actual received care, while satisfaction determinants denote the subjective part where patient characteristics and expectations have an influence. The higher the expectations, the more likely the patients are to be dissatisfied, although some of the

literature argues that the expectation is only a minor part of the actual satisfaction level (Ware et al., 1983).

Previous patient satisfaction studies have in general reported high satisfaction with healthcare (Williams, 1994). This has led to questioning whether it is worth measuring at all. However, more differences and dissatisfaction are elicited when specific components of the GP practice are investigated more carefully (Chow et al., 2009; Vuori, 1991). It seems that having information about the physician before choosing one, and the patient-physician relationship is associated with high patient satisfaction. This was highlighted by Wolinsky and Steiber (1982), who reported that patients considered the physicians reputation in their network, and satisfaction with the GPs interpersonal skills as the most important factors when choosing a doctor. Baker (1990) found three important categories related to general satisfaction: the degree to which the patients trusted the doctors advice, the level of relevant information given and the perception of being given sufficient time. Williams and Calnan (1991) reported from a study in England high scores on communication and professional skills and lower satisfaction with trust, information and consultation length. Related, Gandhi et.al (1997) found from a qualitative study that the main reasons why people changed their physicians in England was accessibility and attitudinal problems concerning the physician being rude and lacking interest during the consultation. Scott and Vick (1999) found that satisfaction is dependent on communication, thorough explanation, quality of information and waiting times. Also McGlone, Butler, and McGlone (2002) identified the importance of the patient-physician relationship, the physicians' ability to communicate and understand, and the time devoted to each patient, as the most important factors for satisfaction. Even though people report various practice characteristics as influencing their choice of GP, evidence suggest that patient satisfaction is the most influenced by the doctor-patient relationship (Pascoe, 1983; Scott, 2000). Bornstein, Marcus, and Cassidy (2000) have reported slightly different results as they found the most important reasons for choosing and changing doctors relate to instrumental medical skills and office management such as waiting time. Physician characteristics such as age and gender contributed the least. An important weakness of this study is that aspects of the interpersonal relationship were not included as a choice, which might explain the results. The effects of competition and whether it drives higher quality and satisfaction have also been studied. As stated above, competition can increase quality

in a price-regulated market (Gaynor 2006). Pike (2010) studied both clinical quality and patient satisfaction in England and found that GPs facing higher competition within a 500 meters radius had increased measures on both of the outcomes.

Patient demographics are a part of the determinants for satisfaction, but the literature is not entirely consistent on what type of characteristics that influence satisfaction the most (Carr-Hill, 1992). Patient age, sex, ethnicity and socio- demographic status have been of particular interest with various results. Age has shown the most consistency, where the elderly are more satisfied than the young (Hall & Dornan, 1990; Pascoe, 1983; Sivertsen, 2014; Williams & Calnan, 1991). Differences between the genders are more ambiguous, where males are found to be more satisfied than females (Williams & Calnan, 1991), females more satisfied than males (Pascoe, 1983), and no difference between genders at all (Hall & Dornan, 1990). Sivertsen (2014) reports mixed results where males are more satisfied only with the GP dimensions regarding availability. Literature regarding GP characteristics can also be found. Lurås (2004) reported that people seemed to prefer older GPs during the implementation of the general practitioner scheme. More recently, Gravelle et al. (2013) argued that the preference for GP is moving towards female doctors and younger aged doctors. The latter is supported by Sivertsen (2014). Furthermore, satisfaction might be influenced by the combination of patient and GP characteristics. Godager (2012) did a study on revealed preferences based on the introduction of the general practitioner scheme. Based on his findings he argues that patients in general prefer a GP who possesses similar observable characteristics, such as age and gender, as them selves. The reasoning behind this is that finding a GP that matches relieves some of the possible agency imperfections. It is likely that communication in the consultation works better regarding diagnosis and treatment decisions when the patient and GP resemble one another. Similar findings is also reported in Sivertsen (2014), based on the same data as the current study, that patients who have GPs that resemble themselves are more satisfied.

1.4.2 Patient satisfaction and its association with patient shortage

Satisfaction studies relating to patient shortage have not been common internationally, thus the primary literature is from the recent years in the Norwegian context. In Lurås (2007) the Norwegian survey of living conditions measuring peoples satisfaction with their GP was applied. The results suggest that patients

enlisted with GPs experiencing patient shortage were less satisfied with the GPs interpersonal skills, medical skills, referral skills and consultation length. Also, the same GPs were measured with higher satisfaction levels on waiting time, although this had little effect on the general satisfaction level. These results indicate that patient shortage is associated with the GPs personality and practice style, which, is supported by Godager and Iversen (2010) who also analyzed the Norwegian survey of living conditions. In general, they found that patient satisfaction has decreased over the time period from 2002 to 2008 in relation to the patient-GP relationship, while satisfaction with booking time, the time between contact and consultation, has increased. The respondents listed with a GP experiencing patient shortage had lower satisfaction levels on issues regarding the patient-doctor interpersonal relationship, than responders with a GP with a full list. This difference remained in the follow up study (Godager & Iversen, 2014). In the period of 2002 to 2008, Godager and Iversen (2010) registered both an actual reduction in booking time, and higher satisfaction level with booking time with the GPs facing patient shortage, compared to the unconstrained GPs. The registered difference with booking time has started to level out between the two groups of GPs, which might be a result of a decreasing proportion of GPs with patient shortage. In the following study, there were no longer a statistically significant difference in booking time between the two GP groups, which also is reflected by the satisfaction levels (Godager & Iversen, 2014).

In her study, Lurås (2007), expected that GPs would shorten the waiting times and increase the length of consultation when faced with patient shortage, based on earlier results by Iversen and Lurås (2002). The fact that satisfaction with consultation length was low might indicate that consultation length is associated with a good patient-doctor relationship, and that this relationship develops over time. This supports the early finding in Morrell, Evans, Morris, and Roland (1986) that satisfaction with consultation time is a GP dimension linked with physician-patient relationship. Howie et.al (1991) report that satisfaction with long consultations was related to the doctor being able to recognize and follow up on the patients needs and suggests that consultation length could serve as a proxy for quality of care. Bjerrum and Sørensen (1992) found that dissatisfaction with the quality of the GP was related to poor communication and being given too little time. All of these findings have in common that the length of the consultation is related to satisfaction, but it is the

interpersonal quality that drives the length, and is not under direct influence by the GP.

These results of patient satisfaction having an association with patient shortage contribute to the hypothesis that patient shortage is an indicator of the doctor having inferior characteristics from the patients point of view (Iversen & Lurås, 2002). This hypothesis gained some further confirmation when Iversen and Lurås (2011) studied the related topic of GP switching. Patients have a perception of the GPs quality level. They consider whether the GP has inferior characteristics when choosing a GP, which is why some GPs are more likely to have patient shortage, which in turn can predict the level of expected switching activity in the future. They concluded from their results in which they found a 50% higher chance of switching from GPs with patient shortage, that patient shortage might be an indicator of inferior GP characteristics and practice style. Iversen and Lurås (2011) suggest that since patient shortage and GP switching reveals patients preferences for the GP, the level of switching activity the GP experiences, should be made public. Another observation they made worth mentioning, is that the factors a GP may apply to attract patients, such as shorter waiting times, high intensive service and more frequent consultations, barely affects the decision of switching GP. Thus it is the satisfaction with the interpersonal relationship that has the higher association with general satisfaction and decision to stay or switch. Iversen and Lurås (2011) also found that those who were allocated to their first choice GPs were less likely to switch at a later time. Thus the more information the patient has when choosing the GP, the higher satisfaction level and lower chance of switching we can expect.

1.5 Study setting

This study is based on data collected from a private website rating GPs, which up to now was unconventional in this context in Norway. In May 26th, 2012 the website Legelisten.no was launched. On this website people can submit anonymous ratings of their current or previous GP along several dimensions describing the GP practice like service, availability and interpersonal relationship. The ratings are made public for other people. This type of ratings has been more common in England and USA, and is important to study as it is shown that a large proportion consumers assess the online ratings when they choose a healthcare provider (Bardach et.al, 2013) and because future support for such sites probably depends on whether they have an

association with systematic surveys and quality indicators (Greaves et al., 2012). Literature on commercial websites where patients rate general physicians or hospitals provide empirical evidence supporting the claim that patient satisfaction in such self-selected samples are correlated with a moderate to strong effect with randomized sampled surveys and clinical quality indicators (Bardach et al., 2013). They have a similar experience in England reported by Greaves et al. (2012). They found a relationship between the hospitals objective measures of quality and the ratings on the NHS choice website. Legelisten is the first commercial rating website on healthcare in Norway, and by conducting studies based on the data material from this site we might gain insight in the usefulness of the rating site and how satisfied people are with their GP and on which dimensions does the satisfaction differ. It has been said that groups of patients are smarter than the individual, and not systematically wrong on information concerning quality (Surowiecki, 2004).

1.6 Aim of the study

The main objective of this study is to investigate whether the data applied from Legelisten can confirm and contribute to already existing results generated from the Norwegian survey of living conditions. More specifically, we will investigate the association between patient satisfaction on selected dimensions of the GPs practice and patient shortage. We will address this relationship by investigating the following two hypotheses: Based on the findings in the satisfaction studies by Lurås (2007), (Godager & Iversen, 2010, 2014) and (Iversen & Lurås, 2011) we formulate the first hypothesis:

H1: Patient shortage is associated with lower satisfaction with the GP-patient relationship.

Based on theory and research findings regarding physician behavior facing patient shortage (Godager & Lurås, 2009; Iversen & Lurås, 2000, 2002) and satisfaction studies (Godager & Iversen, 2010, 2014; Lurås, 2007), we formulate a second hypothesis:

H2: Patient shortage is associated with higher satisfaction with the GPs waiting times.

2 Methodology

2.1 Data

We conduct our empirical analysis on data retrieved from Legelisten.no in March 2013. The dataset consists of anonymous user evaluations of the GP and associated practice. This data is linked with observable characteristics of the GP obtained from the Norwegian Labour Welfare Administration (NAV). Additionally we obtained data from Statistics Norway (SSB) covering age and gender of the Norwegian population. Considering the entire dataset is retrieved from public sites consisting of anonymous evaluations, this thesis is not notified to the Norwegian Social Science Data Service (NSD).

The following analysis is based on a retrospective cross sectional study. We have not planned or developed the measured variables, thus our analysis is limited to the current variables measured on the website. To enable a certain quality level on the evaluations, every submission goes through several screenings, including a qualitative one. These screenings ensure written explanations do not violate any rules concerning accusations and language. Other technical approaches minimize the chance of false submissions. For this thesis, we will only apply published evaluations for the analyses, because these are considered the most valid ones.

2.1.1 Reliability and validity

Reliability denotes the extent of stability in the measurement, e.g. if the same responder surveyed twice produces a consistent result (Sitzia, 1999). The nature of our study limits the possibilities to do such testing. Instead we assess the reliability on a general level. Donabedian (1966) argues it is common when measuring something complex as quality and satisfaction by surveys, to assess reliability by agreement among experts. However, little has been written about such assessment of self-selected surveys. The use of standardized criteria in the survey serves reliability to an extent. We can think of the users as the experts judging the doctors performance and reliability depends on the level of agreement among those rating the same GP. The difference is that the experts are decentralized and have various background and training, which may result in bias. But as Donabedian (1966) points out, when studying complex concepts bias should be acknowledged and accounted for, instead of treating it as non-existing.

Carr-Hill (1992) describes four parameters known to influence response in surveys: choice of population, timing of survey, type of questionnaire, and the measurement of satisfaction. A description of each of the parameters and how they relate to the current study follows. Different populations can have different expectations towards the healthcare. Although expectations do play a role, Pascoe (1983) has found that this does not lead to bias in surveys. It has also been common to consider different social psychological artefacts, which can bias the results such as social desirability (Sitzia & Wood, 1997). However, since the responders are anonymous and it is not intended to measure satisfaction with a specific treatment, the risk of such biases are small (Hekkert et al., 2009). Second, measurements should be done shortly after the encounter to avoid bias in recollection, changes in perception and overlooking issues. In this study, the sample is self-selected and we do not know how far into the past the reviews are. Something internal to the respondent may have changed, and relevant pieces of the story forgotten. The patient viewpoint may also be distorted due to the type of questionnaire. Closed questions provide us with quantitative answers for specific services. Open-ended questions allow the patient to freely comment on what they view as important. The ideal is to use both question types in a survey measuring satisfaction (Chow et al., 2009). This is especially because quantitative ratings tend to be more positive, while open ended questions tend to generate more negative responses (Aharony & Strasser, 1993). The website uses both types of questions, which improves the quality of data. Everybody can view how many stars the GP has received on a specific service, along with a general qualitative description. Only the quantitative responses will be analyzed in this study, but a study comparing the reliability between the quantitative and qualitative description could be conducted in further analyses. The fourth and final parameter described by Carr-Hill (1992) refers to measurement. The measurement of satisfaction is executed by asking the respondents to rank their GP on a numerical 5-point Likert scale. Categorizing performance like this has been one of the standard types of quality measurement (Donabedian, 1966) and widely used in satisfaction studies (Pascoe, 1983). Numerical measurements are handled in the same way, as the traditionally Likert scale ranged from strongly disagree to strongly agree. Ranging the satisfaction level in this way is an attempt to create an interval scale. Often, and also in this case, we cannot assume that the distance between each points is equal, thus it is not a true interval scale, but an ordinal scale. This type of measurement puts constraints on what types of statistical treatments we can apply (Meltzoff, 1997).

We need to consider whether our data is valid. More specifically *"whether operationalization and the scoring of cases adequately reflect the concepts the researcher seeks to measure"* (Adcock, 2001, p. 529). The website seeks to capture patients satisfaction with the GP service. The question is then if these measurements are adequate in describing satisfaction, or if there is something unobservable confounding the results (Sitzia, 1999). The content validity in this study is considered acceptable as the measured indicators are consistent with other similar research capturing satisfaction with GP service (Sitzia, 1999). The selected indicators for satisfaction are based on international websites and correspond closely to the three components Donabedian (1988) discusses of how to evaluate quality in health care. We can also report relative consistency with the measurement variables and results from the satisfaction studies Lurås (2007) and Godager and Iversen (2010) from the Norwegian survey of living conditions.

Whether our findings reflect actual differences in the population, is a question of external validity and generalizability. One of the most important criterias that must be fulfilled to be able to generalize our findings is to draw a representative sample with random sampling. Our study sample is a self-selected sample across Norway. Because of this, the sample has not been stratified to reflect the composition of age and gender, and there might be bias in the sample such as unobserved socioeconomic status (Winter, 2000). The study sample does not adequately resemble the population when we compare the information we have about observable characteristics to generalize across the population. Figure one and two below reveal that young people below 20 and over 60 are clearly underrepresented. Also, the sample includes more women than men, while the majority of age is between 20 and 50 years old. As a result our findings in this paper describe satisfaction for the users of Legelisten at this point.

Figure 1. Age and gender for the study sample (N=14 225)

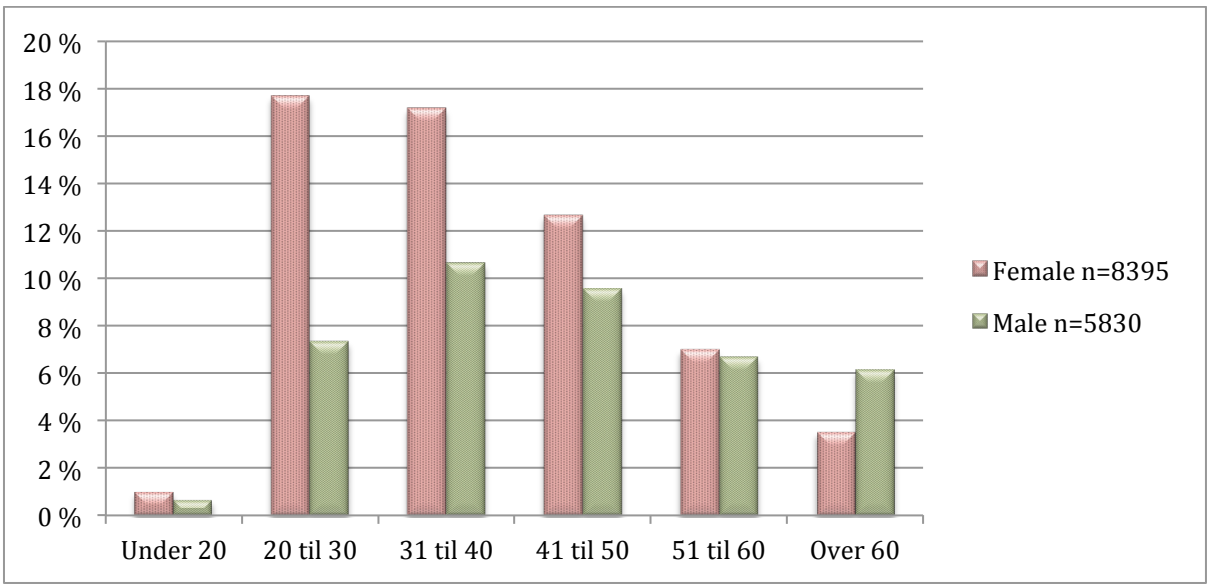
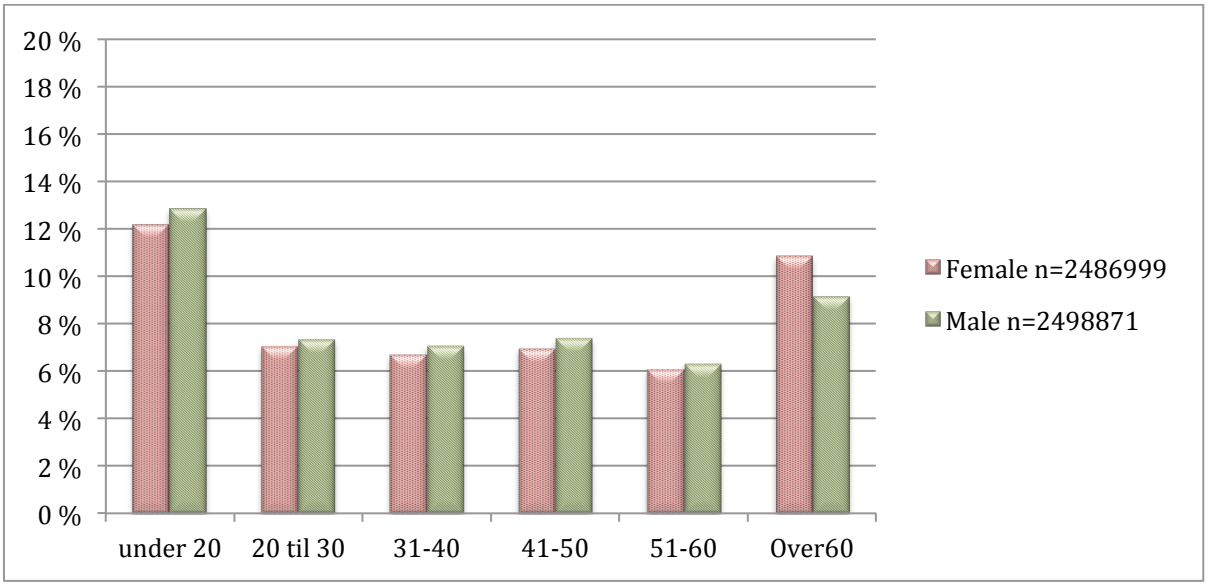


Figure 2. Age and gender for the population obtained from SSB 2012 (N=4 985 870)



2.1.2 Data limitations

The sample is not a result of a well thought sampling strategy, thus is under the risk of certain pitfalls of self-selection such as overrepresentation of groups of people with a certain opinion they wish to share regarding their GP. Additionally, people who had already learned about this website at the point of data extraction, and were inclined to submit an evaluation, may share something unobserved to us. Biased estimators will result if this unobservable heterogeneity is correlated with the explanatory variables in the regression models. Regarding completeness of the dataset, there is a high number of missing values since age and gender are not mandatory to answer. In addition, the number of responses differs between the

dependent variables since it is only mandatory to answer the rating describing overall satisfaction. Data was retrieved when the site had been up and running for less than a year. There is a chance it suffered from currently immature guidelines for how to evaluate, and how to handle the evaluations. Only published evaluations are used in this thesis as these have obliged the rules for evaluation. This is done considering the quality of the data, but it may generate bias in aggregated satisfaction. The non-published evaluations have a higher proportion of dissatisfied patients than the included sample.

2.1.3 Variables

Table 1. Definition of the variables

Variables	Description
Dependent variables	
Overall	How satisfied the responder is summarized over all dimensions from 1 to 5
Booking	How satisfied the responder is with time it takes from contact to the consult takes place from 1 to 5
Waiting	How satisfied the responder is with waiting time in the waiting room from 1 to 5
Consultation	How satisfied the responder is with whether the GP gives them sufficient time in the consult from 1 to 5
Listening	How satisfied the responder is with the GP listening skills from 1 to 5
Insight	How satisfied the responder is with the GPs effort to provide them with insight of own health and treatment from 1 to 5
Advice	How satisfied the responder is with GPs ability to provide trustworthy advice and recommendations from 1 to 5
Independent variables	
Patient shortage	Coded 1 when available list spots>100, coded 0 otherwise denoted as full list
Herfindahl-Hircman index (HHI)	Measures level of GP competition within 10 km radius. Sum of the squares of market share range between 0 and 1 where 1=monopoly
Log Herfindahl (log HHI)	Log of HHI ranging from approximately -6 to 0
GP female	Coded 1 if GP is female, 0 for male
GP age	Continuous variable categorized into dummy variables for age groups, 41-50, 51-60 and over 60. 40 or younger are omitted as reference group
Patient female	Coded 1 if patient is female, 0 for male
Patient age	Dummy variables for age groups 31-40, 41-50, 51-60, over 60. 30 or younger are omitted as reference group
County	Dummy variables for counties
Multilevel Clusters	
GP	Individual GP identification
Zip	Zip code of the GP practice

2.2 Method

This study investigates the relationship between patient satisfaction and patient shortage, controlled for influence of competition, GP and patient characteristics. The association will be investigated with descriptive statistics and multilevel ordered logistic regression analysis. The analysis is conducted with Stata version 12 and Excel version 2011 software tools.

2.2.1 Descriptive statistics

Descriptive statistics is presented with figures and tables in order to demonstrate the immediate relationship between the dependent and independent variables. Mean differences will be tested for significance by Wilcoxon rank sum test. The Wilcoxon rank sum test is the non-parametric alternative to the two-sample t-test used when we cannot assume normal distribution in the bivariate sample. The test is therefore suitable for a logistic distribution where the only assumptions are independent random sample and that the distribution between the bivariate is equal. The test estimates whether the ranked difference in means are statistically significantly different from zero. It is less sensitive to outliers and more robust in estimating the significance level, compared to the t-test (Newbold, Carlson, & Thorne, 2013).

2.2.2 Multilevel ordinal logistic regression

The data consists of dependent variables, of ordered categories from one to five where five is the highest score. The cut points between these categories are known, but assumptions of a certain interval between the categories would be arbitrary since the difference between category 3 and 4 can be different from 4 and 5 (McCullagh, 1980). Categorical variables violate the assumptions for ordinary least square (OLS) regression, therefore applying an ordinal logistic regression model estimated by means of maximum likelihood estimation (MLE) is suitable. Then we avoid restrictive assumptions about linearity and distance between the ordered categories. The MLE method calculates the coefficient estimates that maximize the probability of an outcome that is suitable for the observed data (Long & Freese, 2006; McCullagh, 1980; McKelvey & Zavoina, 1975). A drawback of the MLE approach is that the estimated coefficients is not as easily interpretable as with coefficients estimated with OLS (Long & Freese, 2006).

We model the ordinal responses with a latent variable model. The ordinal regression can be derived by assuming an unobservable latent continuous response variable

underneath the categories. It follows a cumulative logistic distribution F with an assumed variance $(\epsilon) = \pi^2/3$. The categories are then transformed into a set of dichotomies, which represents a series of thresholds values upon the latent variable ranging from $\gamma_0 = -\infty$, and $\gamma_j = \infty$. The thresholds are further used to estimate the cumulative probabilities of a respondent replying one of the categories *s* (Hedeker & Gibbons, 1994; Long, 1997). The ordinal model can be defined by linking the cumulative probability $\Pr(y_i \leq \alpha_s)$ to the linear predictor (McCullagh, 1980; Skrondal & Rabe-Hesketh, 2004,p.29)

$$g[\Pr(y_i \leq \alpha_s)] = k_s - \nu_{I}, \qquad s=1,\ldots,s-1, \tag{3}$$

where $\alpha_1 < \alpha_2 < \cdots < \alpha_s$ are ordered response categories, $\Pr(y_i \leq \alpha_s)=1$ and k_s are unknown thresholds parameters, $k_1 < k_2 < ...k_s - 1$, which do not vary between the units. The probability of a particular response y_s becomes

$$\Pr(y_i = \alpha_s) = \Pr(y_i \leq \alpha_s) - \Pr(y_i \leq \alpha_s - 1). \tag{4}$$

2.2.2.1 Three level structure

We wish to avoid restrictive assumptions of conducting a singular level analysis, which ignores the fact that the individual responses are not independent of each other as some of the GPs have many ratings and a proportion of the GPs have higher rating than others. The respondents are clustered together by the GP, and their market area clusters the GPs, as GPs in the same area are more likely to share the same practice style. Therefore a less restrictive multilevel model, where correlated error terms within levels are assumed, is preferable. Multiple level modeling has been conducted in other studies (Godager et al., 2012; Godager, Iversen, & Ma, 2009; Iversen & Lurås, 2002). In the recent years, it has become more common to include multiple levels in the models as the development of proper tools has advanced. Groupings or clusters of responses can produce more valid results rather than just accounting for level one individual response (Goldstein, 2011). Respondents belonging to the same clusters, share cluster-specific influences, thus they are correlated. This is accounted for by including GPs and zip code as level two and three in the analysis. Then we are able to explain more of the unobserved heterogeneity influencing the responses. However, it is not realistic to observe or include all possible relevant cluster specific covariates in the analysis. As a result we can expect there to be unobserved heterogeneity not only between clusters, but also within the cluster levels after conditioning on the covariates. Therefore we model random effects that are allowed to have different variance between levels. A model

can include random intercept, coefficients, or both. For our purpose we will include random intercepts, which represent unobserved heterogeneity in the overall response (Skrondal & Rabe-Hesketh, 2004). The model can be specified as:

$$y^*_{ijk} = \beta_2 x_{2ijk} + \cdots + \beta_n x_{nk} + \delta_{jk}^{(2)} + \delta_k^{(3)} + \epsilon_{ijk} \quad (5)$$

where y^*_{ijk} is the latent unobserved continuous response variable, which is related to the observed ordinal variable y_{ijk} via the threshold model described above. χ_{ijk} is a covariate with fixed regression coefficients β, $\delta_{jk}^{(2)}$ and $\delta_k^{(3)}$ are random intercepts at level two and three respectively, and ϵ_{ijk} is the unobserved variance following a logistic distribution, independently of x_{ijk}, $\delta_{jk}^{(2),}$ and $\delta_k^{(3)}$ (Rabe-Hesketh & Skrondal, 2012).

2.2.2.2 Assumptions

Logistic regressions require assessing some underlying assumptions. When assumptions are not met in logistic regression, we risk reporting biased and unstable coefficients and standard errors, which in turn can lead to invalid statistical inferences (Midi, Sarkar, & Rana, 2010). Logistic regression does not require the assumptions of OLS such as normal distribution of error terms, linearity and homoscedasticity due to the non-linear transformation of the outcome (O`Connell, 2006).

The main assumption for using ordinal logistic regression is the proportional odds assumption. This assumption states that slopes between categories must be constant, which means the odds of being in a higher category must be stable (O`Connell, 2006). The proportional odds assumption is testable in Stata with the brant test. Our models did not fulfill the assumption, which could be expected as our models consist of a large sample size, many explanatory variables and one of the independent variables is continuous. The violation indicates that some of the variables contain odds between categories that are significantly different from each other (O`Connell, 2006). However proportional odds tests are very sensitive, often violated and should be interpreted with caution (Allison, 1999). One possible solution is to relax the alpha level. Another solution is to use a different analysis, but the key approach is to identify which of the variables that is responsible for the violation, and then investigate the magnitude of the inconsistent slopes (O`Connell, 2006). The gender variables and a couple of the age groups were the source of violation in the models. Since the direction of the slope was constant between categories, and these variables

serve merely as controlling variables in the models, the violation of the assumption did not generate great concern. In addition, it has been recommended to evaluate variation in effects with a binary outcome, which is included in the appendix (Brant, 1990; O`Connell, 2006). Additional assumptions exist for logistic regression described by Aldrich and Nelson (1984). Logistic regression assumes linearity between the independent variable and the logit, which cannot be assessed directly in ordinal models. If the assumption is incorrect, we risk understating the association. Next it is important to handle binary and ordinary outcomes properly. Collapsing an ordinal outcome into a binary involves loosing a lot of information and risk bias in the results. Although it might be tempting to simplify the analysis by collapsing the categories, we have kept our five-category outcome, but presents a binary alternative in the appendix. Further, the observations in the dependent variable are assumed to be independent of one another, and consist of a random sample of units (Aldrich & Nelson, 1984). Although the responses are independent of each other considering this is not a panel or a matched pair study, we may assume there to be some correlation between the responders who share the same GP and GPs within the same geographical areas, which led us to clustering in the model. It is also normal to assume absence of multicollinearity which occurs if there is a strong linear relationship between the independent variables (Aldrich & Nelson, 1984). Midi et al. (2010) explain multicollinearity as a "monster" in regression analysis that is challenging to tame. The best way to avoid it is by fitting the model with caution by stepwise selection, but even then the presence of multicollinearity may corrupt the process. The model as a whole may still serve as a good explanation of the variation in the response variable, thus it is the individual prediction that may become unstable with inflated standard errors. Simple diagnostics were performed such as correlation matrix where none of the correlations exceeded 0.8, and checking the variance inflation factor (VIF). Logistic regression is a less powerful analysis than OLS, thus the VIF value should not exceed 2.5. A few VIF values exceeded this threshold, although the mean VIF is 1.76. If possible one should drop some of these variables, centering the mean, create interaction variables or possibly the best solution, to be aware of its presence and its consequences. The latter will be the approach in this analysis. When we have a large sample size with sufficient observation for each case, multicollinearity is not a major problem (Midi et al., 2010).

2.2.3 The empirical model

"While measures of fit provide some information, it is only partial information that must be assessed within the context of the theory motivating the analysis, past research, and the estimated parameters of the model being considered (Long, 1997, p. 102).

Model fitting was mainly conducted by including variables based on the theoretical and empirical reasoning presented in chapter one, and using the likelihood ratio test, which has been argued to be the most reliable for assessing independent variables' contribution to a model (O`Connell, 2006). The main explanatory variable patient shortage along with the covariates *log herfindahl age* and *gender* of both patients and GPs were included based on the study objective and previous research. Patient shortage proved to be a strong and stable effect in the pre analysis. Next, the dummy variables for *counties* were tested for relevant inclusion based on a possible argument that there is an individual difference that influences satisfaction reflected in the choice of residency in rural or city areas. The inclusion of *county* had a major influence by reducing the initial effect of *log HHI* on satisfaction. We concluded based on the log-likelihood ratio test that inclusion of counties had a significant effect different from zero and contributed to variance in the regression models: *overall, booking, and waiting*. Regression models for *consultation, advice, listening and insight* resulted in an insignificant contribution. However, *county* is still included in the models as the influence, although not as strong, on *log HH*I remained. Likelihood ratio tests were also conducted for inclusion of the third level random effect, *zip*. *Zip* tested to provide a significant contribution to the variance in all regressions at 1 and 5% level with exception for *consultation* and *advice*. *Zip* is included in these models as well as it still proves to capture some unobserved variance and to uphold consistency with the other models.

The specific model analyzing the relationship between the dependent satisfaction variables *overall, booking, waiting, consultation, listening, insight, advice,* and explanatory variables is expressed below.

$$y^*_{ijk} = \beta_1 \text{Patientshortage}_{ijk} + \beta_2 \text{LogHerfindahl}_{ijk} + \beta_3 \text{GPfemale}_{ijk} + \beta_4 \text{GPage}_{ijk} + \beta_5 \text{Patientfemale}_{ijk} + \beta_6 \text{Patientage}_{ijk} + \beta_7 \text{County}_{ijk} + \delta_{jk}^{(2)+} \delta_k^{(3)} + \varepsilon_{ijk}$$

The multilevel ordered logistic regression models are conducted using a user written Stata program, General Linear Latent Mixed Models approach (GLLAMM). This program can estimate regression analysis based on the common type of dependent variables, including ordered categories (Rabe-Hesketh & Skrondal, 2003b).

3 Results

3.1 Descriptive presentation

This part presents descriptive statistics of the general distribution of the dependent variables before restrictions on the sample are imposed. Followingly we check whether there is bias present in the composition of GPs and patients demographics within the grouping of the main explanatory variable *patient shortage*. Lastly we look into differences in means on the dependent variables by *patient shortage* and conduct a significance test.

Responders' ratings on the website Legelisten seem to follow the same pattern documented in earlier research. It is common to find that people are very satisfied with their GP on a general level (Williams et al., 1998).

Figure 3. Distribution of satisfaction on the dependent variables

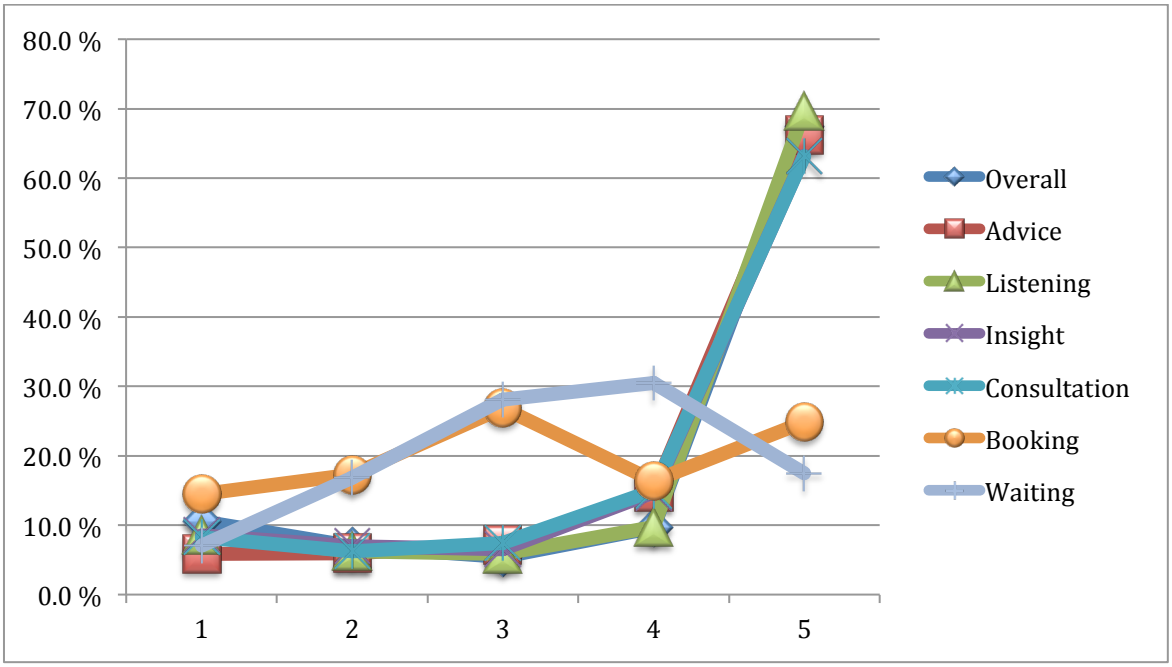


Figure 3 shows that the responders are very satisfied with their GP and the main proportion has rated their GP at the highest category 5. The dependent variable *overall*, and the variables describing the interpersonal relationship; *advice*, *listening*, *insight* and *consultation* have been rated at the highest category by approximately 70% of the responders. High satisfaction with interpersonal relationship has been described as being important in the literature (Baker, 1990; McGlone et al., 2002). Booking and waiting stand out with a larger variance between response categories,

and a much lower proportion, under 30% is very satisfied. This figure reveals that the current sample rates their GP in the same pattern as reported in previous studies, high satisfaction overall and with dimension describing the interpersonal relationship, and lower and wider spread on waiting times (Godager & Iversen, 2014). This is an important observation since the sample is of an unconventional type in the Norwegian context previously predicted to generate mostly negative responses (Danielsen, Kjøllesdal, & Bjertnæs, 2013).

Earlier research has reported varying results regarding whether satisfaction is dependent on demographics such as age and gender (Carr-Hill, 1992). Sivertsen (2014) has analyzed the difference in satisfaction dependent on demographics based on the current dataset. Satisfaction increase with patient age, and male patients are more satisfied than female patients with *booking* and *waiting*. Responders are also more satisfied if the GP is male, and if the GP is of younger age. Since demographics seem to influence satisfaction, we present the composition of GPs and patient characteristics between the GPs with *patient shortage* and full list. This is done in Figure 4 and 5 below in order to investigate whether there are any demographic biases that could influence the satisfaction levels between the groups.

Figure 4. GP demographics between patient shortage and full list

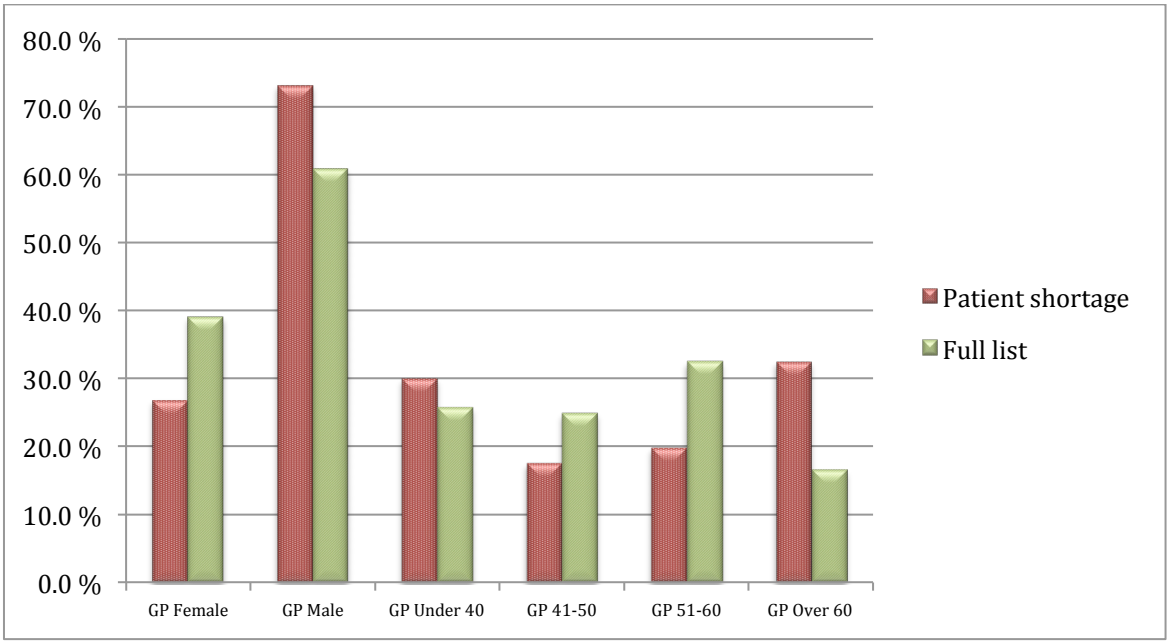


Figure 5. Patient demographics between patient shortage and full list

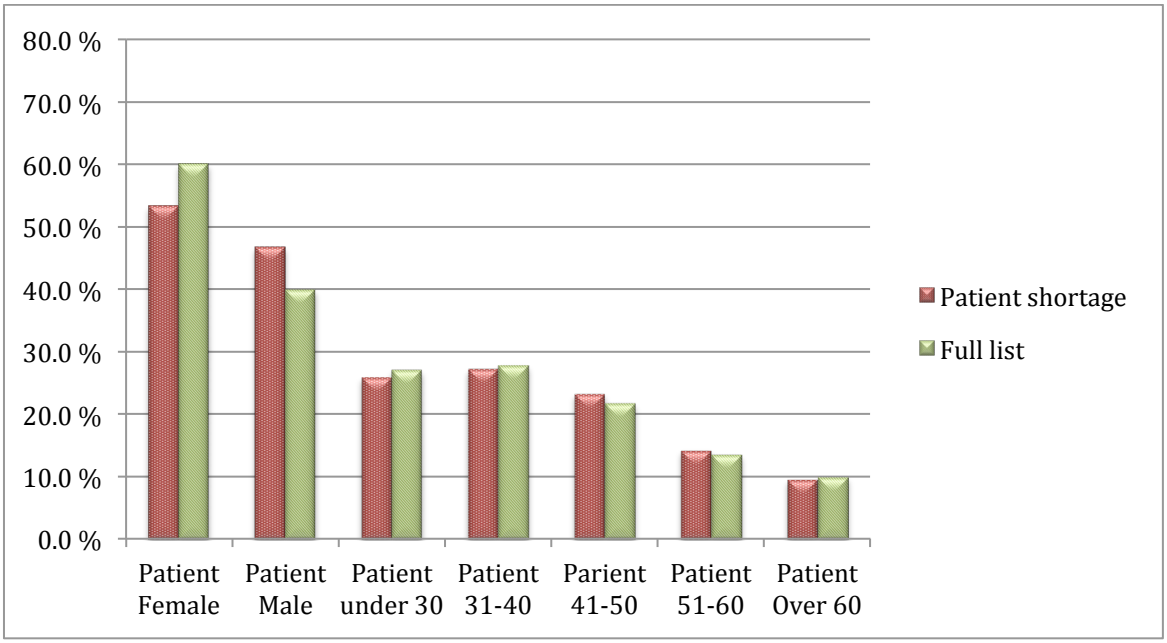


Figure 4 shows that it is a larger proportion of male GPs with *patient shortage* and a slightly larger proportion of *patient shortage* among the youngest and oldest GPs.
Figure 5 shows the patient demographics. There are no noteworthy differences in the composition of patient demographics between *patient shortage* and full list except that males are slightly underrepresented in the full list group compared to the *patient shortage* group.

Table 2 presented below shows whether there is a difference in mean satisfaction on the dependent variables between *patient shortage* and full list. It shows that the mean satisfaction is in accordance with previous literature where satisfaction is found to be lower on dimensions describing the interpersonal relationship and higher on waiting times, when the GP has patient shortage (Godager & Iversen, 2014; Lurås, 2007).

Table 2. Mean satisfactions divided by patient shortage and full list

	Patient Shortage			Full list			
Variable	**N**	**Mean**	**SE**	**N**	**Mean**	**SE**	**Prob>|z|**
Overall	2042	3.939	1.529	15924	4.195	1.366	0.0000
Booking	1554	3.377	1.313	12639	3.177	1.373	0.0000
Waiting	1560	3.428	1.199	12717	3.335	1.15	0.0008
Consultation	1637	4.078	1.379	13195	4.196	1.279	0.0109
Advice	1645	4.122	1.306	13248	4.319	1.169	0.0000
Listening	1649	4.028	1.446	13269	4.286	1.284	0.0000
Insight	1624	3.988	1.423	13120	4.194	1.293	0.0000

Table 2 shows the mean and standard error of satisfaction on the dependent variables between responders rating a GP with *patient shortage*, and full list. We can see that mean satisfaction is lower for the *patient shortage* group on *overall*, *consultation*, *advice*, *listening* and *insight*. On the other hand, the mean satisfaction is higher for the same group on *booking* and *waiting*. *Patient shortage* and full list tested to be significantly different from each other at a 1% significant level on all dependent variables with the non- parametric Wilcoxon rank sum test (Mann Whitney).

3.2 Regression analysis

Below are three tables presented with regression results for the models with the dependent variables: *overall, advice, listening, insight consultation, booking and waiting* estimated from the following model; Multilevel ordered logistic regression

$$y^*_{ijk} = \beta_1 \text{Patientshortage}_{ijk} + \beta_2 \text{Logherfindahl}_{ijk} + \beta_3 \text{GPfemale}_{ijk} + \beta_4 \text{GPage}_{ijk} + \beta_5 \text{Patientfemale}_{ijk} + \beta_6 \text{Patientage}_{ijk} + \beta_7 \text{County}_{ijk} + \delta_{jk}^{(2)} + \delta_{k}^{(3)} + \varepsilon_{ijk}$$

The primary emphasis for these regressions is to estimate the possible influence of *patient shortage* on the dependent satisfaction variables, conditioned by the covariates. In general, the regression results support the findings in the descriptive analysis. All of the models have estimated associations between patient shortage and patient satisfaction as predicted in the hypothesis. These results are aligned with previous literature, which is discussed in more detail in the discussion section.

Table 3. Multilevel ordered logistic regression for overall

Model 1: Overall		
Predictors	**exp (b)**	**95% Confidence interval**
Patient shortage	0.596**	(0.489, 0.728)
Log herfindahl	1.008	(0.941, 1.079)
GP Female	0.816**	(0.713, 0.935)
GP age 41-50	0.743**	(0.619, 0.892)
GP age 51-60	0.595**	(0.501, 0.708)
GP age 60+	0.524**	(0.429, 0.642)
Patient Female	1.041	(0.942, 1.149)
Patient age 31-40	1.288**	(1.138, 1.457)
Patient age 41-50	1.555**	(1.360, 1.778)
Patient age 51-60	1.993**	(1.695, 2.344)
Patient age 60+	2.352**	(1.950, 2.837)
County dummies ±	Yes	(0.851, 1.670)
K1	-3.116**	(-3.507, -2.725)
K2	-2.405**	(-2.793, -2.018)
K3	-1.971**	(-2.357, -1.586)
K4	-1.261**	(-1.644, -0,877)
Random Effect	**Variance**	**Standard error**
$\boldsymbol{\delta_{jk}^{(2)}}$ GP	1.1089	0.945
$\boldsymbol{\delta_{k}^{(3)}}$ Zip	0.0743	0.0386
Log Likelihood	-10559	-
NO Response	11262	-
NO GPs	3250	-
NO Zip	893	-

Significant different from zero at **1% level and *5% level.
± non are significant at a 5% level.

Results from model 1 support the findings from the descriptive analysis. We see that *patient shortage* has a negative association with *overall*. Patients enlisted with a GP facing *patient shortage* have lower odds of 0.596 for high satisfaction compared to *full list*, when all other covariates are held constant. Further we see that both GP and patient characteristics influence the satisfaction level. The odds for high satisfaction are 0.816 lower for *GP female* compered to *GP male*. Also, higher *GP age* has a negative association with satisfaction levels compared to GPs who are 40 or younger. Among the patient characteristics, only the *patient age* groups have a significant influence on satisfaction levels. The odds of high satisfaction increase with higher age compared to those who are 30 or younger. A higher *log Herfindahl* indicates a higher market concentration, which proxies less competition. After the inclusion of the *county* dummies, satisfaction *overall* is not significantly associated with competition level.

The estimated threshold values *k1, k2, k3* and *k4* are shown at the bottom of the fixed effect part of the table. These thresholds mark the cut point where it is predicted to rate the GP in a higher category, and is normally not given an individual interpretation (O`Connell, 2006). Further we have estimated two random intercepts. The variance between *GPs* is 1.108 and the variance between *zip* codes is .074. These random intercepts show how much heterogeneity there is between the individual GPs and zip code in model 1, conditioned by the fixed covariates.

Table 4. Multilevel ordered logistic regression for time use

	Model 2:Booking		**Model 3: Waiting**		**Model4: Consultation**	
Predictors	**exp (b)**	**95% CI**	**exp (b)**	**95% CI**	**exp(b)**	**95% CI**
Patient Shortage	1.389**	(1.180, 1.635)	1.209*	(0.996, 1.468)	0.752**	(0.620, 0.911)
Log herfindahl	0.922**	(0.869, 0.979)	0.945	(0.883, 1.013)	1.033	(0.969, 1.101)
GP female	0.703**	(0.633, 0.781)	0.766**	(0.677, 0.867)	0.844**	(0.743, 0.959)
GP 41-50	0.897	(0.782, 1.030)	0.793**	(0.673, 0.934)	0.773**	(0.651, 0.917)
GP 51-60	0.886	(0.777, 1.012)	0.603**	(0.516, 0.705)	0.622**	(0.529, 0.731)
GP 60+	0.860	(0.734, 1.006)	0.575**	(0.477, 0.963)	0.641**	(0.529, 0.776)
Patient female	0.873**	(0.806, 0.944)	0.835**	(0.769, 0.907)	1.129**	(1.028, 1.239)
P 31-40	1.102	(0.995, 1.220)	1.052	(0.946, 1.169)	1.242**	(1.104, 1.398)
P 41-50	1.077	(0.966, 1.201)	1.220**	(1.089, 1.366)	1.546**	(1.359, 1.758)
P 51-60	0.958	(0.846, 1.085)	1.485**	(1.303, 1.692)	1.559**	(1.342, 1.810)
P 60+	0.981	(0.852, 1.131)	1.677**	(1.444, 1.947)	2.153**	(1.805, 2.659)
County±	yes	-	yes	-	yes	-
k1	-2.339**	(-2.675, -2.004)	-3.276**	(-3.66, -2.886)	-3.362**	(-3.735, -2.989)
k2	-1.163**	(-1.495, -0.831)	-1.525**	(-1.906, -1.143)	-2.659*	(-3.028, -2.290)
k3	-1.163	(-0.157, 0.505)	0.118	(-0.262, 0.498)	-2.039**	(-2.406, -1.673)
k4	1.053**	(0.721, 1.385)	1.991**	(1.609, 2.373)	-1.099**	(-1.463, -0.753)
Random Effect	**Variance**	**Standard error**	**Variance**	**Standard error**	**Variance**	**Standard error**
$\delta_{jk}^{(2)}$ GPs	0.530	.0544	1.135	.0787	0.996	.083
$\delta_{k}^{(3)}$ Zip	0.236	.038	0.292	.052	0.047	.033
Log Likelihood	-16132	-	- 15053	-	-11402	-
NO Response	10494	-	10556	-	10877	-
NO GPs	3250	-	3188	-	3217	-
NO Zip	893	-	886	-	889	-

Significant different from zero at **1% and *5% level.
± non are significant at a 5% level.

Table 4 shows the results from model 2, 3 and 4. From model 2 with *booking* as dependent variable, we see that *patient shortage* has a positive association with satisfaction levels with *booking*. The odds of high satisfaction when the patient is listed with a GP experiencing *patient shortage* are 1.389 times higher compared to full list, conditioned by the other covariates. We register a significant negative

association between *log Herfindahl* and satisfaction. The odds of high satisfaction when patients have a GP with *patient shortage* are 0.922 lower when *log Herfindal* increases by one percent. This indicates that as the market becomes less competitive, the patients are less likely to be satisfied with *booking*. Regarding the GP characteristics, only the gender variable shows a significant influence where the odds of higher satisfaction are 0.703 lower for *GP female* compared to males. The same goes for patient characteristics, as the odds of satisfaction decrease for *patient female* compared to males. Further we have estimated two random intercepts. The between *GP* variance is .5302 and between *zip* variance is .236. These random intercepts show how much observed variation there is between the individual GPs and zip codes.

From model 3 with *waiting* at the GPs office as the dependent variable, we can report a positive and significant association between *patient shortage* and satisfaction with *waiting*. The odds of high satisfaction are 1.209 higher when the GP faces *patient shortage* compared to those having a full list, conditioned by all other covariates. The effect of *log HHI* shows that satisfaction decrease as the market gets less competitive, but is not significant. The GP characteristics all show a negative association with satisfaction. The odds of high satisfaction are 0.766 lower when the GP is *female* and decrease with higher *GP age* groups compared to GPs who are under 40. The odds of high satisfaction are also lower for *female patients* compared to males and decrease by a significant rate from the *patient age* group 41-50 and up, compared to patients aged 30 or younger. Further we have estimated two random intercepts. The variance between *GPs* is 1.135 and the variance between *zip* codes is .292. These random intercepts show how much observed variation there is between the individual GPs and zip codes. *GP* has the second largest variance in model 3 and *Zip* has the largest variance in models 2 and 3 out of all the estimated models. This can be interpreted to mean that unobserved heterogeneity on the market level is mostly relevant for models measuring satisfaction on dimension that the GP can influence.

In Model 4 with *consultation* as dependent variable we see a negative and significant association between *patient shortage* and satisfaction with *consultation*. Even though consultation denotes time use, it has been shown to reflect the interpersonal relationship and therefore not a subject of influence by the GP (Lurås, 2007; Morrell et al., 1986). When patients are enlisted with a GP facing *patient shortage* the odds of high satisfaction are 0.752 lower compared to full list, conditioned by all other

covariates. The effect of *log HHI* shows that the odds of being satisfied increase as the market get less competitive, but is not significant. Further, there is a negative association between the GP characteristics and satisfaction. The odds are 0.844 lower for *GP female*, and decrease with higher *GP age* compared to when GPs are 40 or younger. Patient characteristics have positive and significant associations with satisfaction. The odds of high satisfaction are 1.29 higher for *patient female* compared to males and increase with higher *patient age* compared to patients 30 or younger. The variance between *GPs* is estimated to .996 and *zip* code is .047 at level two and three respectively. These random intercepts show how much observed variation there is between the *GPs* and *zip*. *Zip* is much lower in the *consultation* model compared to *booking* and *waiting*.

Table 5. Multilevel ordered logistic regression results for interpersonal relationship

	Model 5: Advice		**Model 6: Listening**		**Model 7: Insight**	
Predictors	**exp (b)**	**95% CI**	**exp (b)**	**95% CI**	**exp (b)**	**95% CI**
Patient Shortage	0.657**	(0.548, 0.789)	0.565**	(0.459, 0.695)	0.675**	(0.562, 0.811)
Log herfindahl	1.002	(0.943, 1.066)	1.023	(0.952, 1.099)	1.025	(0.963, 1.091)
GP female	0.862*	(0.762, 0.975)	0.866*	(0.752, 0.988)	0.868*	(0.767, 0.982)
GP 41-50	0.839*	(0.713, 0.988)	0.709**	(0.587, 0.858)	0.813*	(0.691, 0.958)
GP 51-60	0.723**	(0.619, 0.844)	0.572**	(0.477, 0.685)	0.678**	(0.580, 0.793)
GP 60+	0.707**	(0.587, 0.850)	0.508**	(0.412, 0.628)	0.657**	(0.546, 0.790)
Patient female	1.035	(0.941, 1.139)	1.051	(0.948, 1.164)	1.005	(0.916, 1.103)
P 31-40	1.114	(0.988, 1.354)	1.168*	(1.028, 1.326)	1.135*	(1.010, 1.275)
P 41-50	1.432**	(1.257, 1.630)	1.594**	(1.385, 1.835)	1.444**	(1.272, 1.639)
P 51-60	1.498**	(1.287, 1.743)	1.779**	(1.507, 2.100)	1.597**	(1.376, 1.854)
P 60+	2.153**	(1.794, 2.584)	2.186**	(1.800, 2.654)	2.197**	(1.840, 2.624)
County $\pm$	yes	-	yes	-	yes	-
k1	-3.518**	(-3.878, -3.158)	-3.474**	(-3.889, -3.058)	-3.147**	(-3.505, -.2.789)
k2	-2.670**	(-3.023, -2.317)	-2.781**	(-3.192, -2.370)	-2.388**	(-2.741, -2.034)
k3	-2.013**	(-2.362, -1.663)	-2.255**	(-2.663, -1.846)	-1.860**	(-2.211, -1.508)
k4	-1.010**	(-1.357, -0.663)	-1.538**	(-1.944, -1.131)	-0.912**	(-1.261, -0.563)
Random Effect	**Variance**	**Standard error**	**Variance**	**Standard error**	**Variance**	**Standard error**
$\delta_{jk}^{(2)}$ GP	0.782	0.074	1.187	0.102	0.849	0.075
$\delta_{k}^{(3)}$ Zip	0.039	0.029	0.096	0.044	0.055	0.030
Log Likelihood	-10634	-	-11424	-	-9983	-
NO Response	10904	-	10912	-	10831	-
NO GPs	3217	-	3220	-	3214	-
NO Zips	890	-	891	-	891	-

Significant different from zero at **1% level and *5%level.
$\pm$ non are significant at a 5% level.

Table 5 shows the results from model 5, 6 and 7. Model 5 with *advice* as dependent variable shows a negative and significant association between *patient shortage* and satisfaction with *advice.* The odds of high satisfaction are 0.657 times lower when patients are enlisted with a GP facing *patient shortage* compared to full list, conditioned by all other covariates. All GP characteristics show a negative and significant association with satisfaction. The odds of high satisfaction are 0.862 lower for *patient female,* and decrease with higher *GP age* compared to when the GP is 40 or younger. The odds of high satisfaction significantly increase with higher *patient age* from the age group 41-50 and up. The estimated variance at the second level for *GPs* is 0.782, and 0.039 for *zip* at the third level. In this model, the third level inclusion was not significant. However it still explains some unobserved heterogeneity and is included in the model[1].

Model 6 shows a negative and significant association between *patient shortage* and satisfaction with the GPs *listening* skills. The odds of high satisfaction are 0.565 lower when patients are enlisted with a GP facing *patient shortage* compared to patients with a GP with a full list, conditioned by all other covariates. The effect of *log HHi* shows that the odds of being satisfied increase as the market gets less competitive, but is not significant. The GP characteristics have a negative association with satisfaction. The odds of high satisfaction are lower when the rated GP is *female*, and decrease with *GP age* compared to GPs 40 or younger. *Patient age* has a positive association with satisfaction and increases as the patient is older compared to patients who are 30 or younger. The estimated variance at the second level between *GP*s is 1.187, which is the largest value between all the models, and 0.096 for *zip* at the third level.

Model 7 shows a negative and significant association between *patient shortage* and satisfaction with GPs contribution to *insight*. The odds of high satisfaction are 0.675 lower when patients are enlisted with a GP who faces *patient shortage* compared to full list, conditioned by all other variables. The effect of *log HHI* is not significant. All GP characteristics have a negative association with satisfaction. The odds of high satisfaction are lower for *female GPs* and with increasing *GP age*, compared to GPs who are 40 or younger. *Patient age* has a positive significant association where

[1] It should be noted that results are not sensitive to whether or not a random effect at the third level is included.

satisfaction increase as the patient is older, compared to patients who are 30 or younger. The variance between *GPs* is estimated to 0.849 and the variance between *zip* is 0.055.

In order to check for robustness it is conducted binary analysis for the seven ordinal models for comparison presented in the appendix. The dependent variables are coded 1 if the rating=5, and 0 otherwise, hence we have a strict dichotomizing. Comparing the two different models for *overall*, we see very small differences in the odds ratios. The main predictor *patient shortage,* has a significant and slightly weaker effect in the binary regression (0.613) than in the ordinal regression (0.596) on *overall*. A similar tendency is true for all the variables describing the interpersonal relationship between the GP and the patient. There is a larger difference between the models for *booking* and *waiting*. In the binary analysis, the odds ratio for being satisfied with *waiting* when the GP has patient shortage is higher in the binary (1.405) than in the ordinal analysis (1.209). The odds for being satisfied with *booking* are lower and not significant in the binary analysis. The probable reason for this deviance is that this category had a much larger spread over the response categories compared to the variables describing the interpersonal relationship. Overall the binary models confirm the effects from the ordinal model, which provides us with evidence that the analyses are strong and robust. The effects from model 2, *booking* should be interpreted with caution.

4 Discussion and Conclusion

This section discusses the main findings in light of previous literature and points to limitations and strengths of this analysis. It also discusses possible further research and implications that can be inferred from the results. A concluding remark is drawn at the end.

4.1 Main findings

The main objective for this study is to investigate the relationship between patient satisfaction with the GP and patient shortage. To analyze this association we have conducted a multilevel ordinal regression where the main outcome is the odds ratio of a response in a higher category of satisfaction when the GP has patient shortage. Two hypotheses were stated to address the objective. The results in this study support the first hypothesis that *Patient shortage* is associated with lower rated satisfaction with the GP dimensions that describe the interpersonal relationship between the patient and GP. We have seen from both the descriptive and regression analysis that *overall* satisfaction, which captures a summarized rating including all the aspects of the GP practice, is lower for patients rating a GP who has *patient shortage* compared to the GPs with a full list. This finding has shown to be robust, as it has persisted with a strong effect through every type of pre and main analysis. The responses have shown through the descriptive presentation and the regression analysis, lower odds for high satisfaction when the GP has *patient shortage* on four dimensions of the interpersonal relationship. These are the GPs ability to give trustable *advice*, *listen* to what the patient have to say, communicate better *insight* of disease and treatment and the perception of being given sufficient time in the *consultation*. These four dimensions also correspond with *overall*, which indicates that these dimensions have the strongest influence on *overall* satisfaction. The results also support the second hypothesis that *patient shortage* is associated with higher rated satisfaction with the GPs waiting times, which may be influenced by the GP. Both descriptive and regression results confirm the association. Satisfaction with *booking*, the time from initial contact to when the consultation takes place is more likely to be high when the GP faces *patient shortage*. The same is true for satisfaction with *waiting*, the time spent in the waiting room.

The association between patient satisfaction and *patient shortage* in the current study confirms and adds to previous empirical results. Several papers have based their

research on the Norwegian survey of living conditions. Lurås (2007) found that patients enlisted with a GP with patient shortage were less satisfied with the GPs interpersonal skills and with the consultation length. The same patients were on the other hand more satisfied with waiting time, although it did not seem to affect the overall satisfaction level. These results imply that it is the interpersonal dimensions that matter the most. This is confirmed by Iversen and Lurås (2011) who comment that the improvement on availability among GPs with patient shortage as observed in their study, does not seem to compensate for the dimensions considered as inferior. We can argue that this applies in the current study, as it seems that it is the interpersonal relationship that contributes to overall satisfaction, not shorter waiting times. Godager and Iversen (2010); (2014), have found similar results. Patients enlisted with a GP with patient shortage had lower satisfaction on the interpersonal relationship, and higher satisfaction with booking. Satisfaction with booking has shown to reflect the actual changes in booking times, which is interesting because it demonstrates how responsive the patients are to actual change.

The current study has a wider range of variables describing the interpersonal relationship than those described above, which gives interpretation of the association more depth. In addition, unlike the previous studies, this study has measurements on both booking and waiting, which in reality are two different types of waiting. A GP offering a shorter wait for consultation might not necessarily offer a short wait in the waiting room. Our results indicate that patient shortage induce a reduction on both.

Satisfaction is a complex measure that can be influenced by other factors. We therefore included several covariates based on theory and empirical results in order to control for influence on the association of interest. *Log Herfindahl* measures market competition for GPs and seemed at first to influence higher satisfaction with lower competition levels, except for satisfaction with *booking* and *waiting*. However, this association diminished after the inclusion of *county*. This change in effect can be interpreted as that *log HHI* picked up something else in addition to competition, e.g. that patients living in the cities and rural areas have different expectations influencing satisfaction. The competition measure remained as a significant effect in the model for *booking*, where satisfaction is more likely to be lower as the market gets less competitive. *Patient* and *GP age* and *gender* were also included in the models. In

general they proved to affect satisfaction in the same manner as has been reported in earlier studies. Patient satisfaction increases with *patient age*, which is the most consistent demographic influence on satisfaction in the literature (Hall & Dornan, 1990). Patient gender had a significant effect in three of the models. Males have higher odds of being satisfied with *booking* and *waiting*, while *females* have higher odds of being satisfied with the *consultation*. The literature has previously found mixed and contradictive results, (Pascoe, 1983; Williams & Calnan, 1991). The results of gender in this study might have been able to capture on what specific dimensions the difference lies. The current study also finds that satisfaction is more likely to be lower when the GP ages, which is aligned with the findings by Gravelle et al. (2013).

4.2 Limitations and strengths

The covariates that were available and relevant for the objective of the study were included in the models. However, even though this allowed us to control for influence on satisfaction, there might be additional influencing characteristics we have not been able to control for. For instance, since the age categories for patients are ranged widely, it was problematic to include a matching variable consisting of age and gender for patient and GP. Godager (2012) suggests that when patient and GP resemble one another on observable characteristics, there is an increased possibility for higher satisfaction. Such a relationship has been indicated for the applied data (Sivertsen, 2014). The results show that patients are less likely to be satisfied if the GP is *female* and if the GP is of an older *age group*. The problem with this pattern is that the majority of the older GPs are male. Further, there might be more unobservable heterogeneity that we have not been able to account for in this analysis. For example, self-declared health status, education, the continuity of the GP-patient relationship and the number of switches the GP experiences, could have provided some additional explanation. Including these variables and enabling for creating a matching variable could be relevant for further studies.

Another limitation is that although evidence suggests an association, we cannot draw inferences related to the causality of this relationship. There are also some limitations to the interpretation of models of the type used in this analysis. The estimated coefficients are not as intuitively interpretable as with OLS or binary logistic regressions, as the probability of change in one independent variable depends on the levels of all covariates in the model (Long & Freese, 2006). However the binary analysis in the appendix has demonstrated that our estimated odds ratios

are relatively robust, since most of the effects remained the same. There is one difference in the binary analysis that must be mentioned; *patient shortage* does not have a significant effect on satisfaction with *booking*. This might lead us to question how much we can rely on this estimate. The probable cause is the strict choice of cutoff point between the ordinal response categories to create a binary outcome and that satisfaction with *booking* has the widest distribution of all the dependent variables.

It is possible to conduct further post analysis on logistic and ordinal logistic regression. Such analysis has not been conducted for this study as the multilevel structure makes these analyses more complicated and less informative. On the other hand, the multilevel structure makes it possible to capture additional unobserved heterogeneity by clustering responders by GPs, which are further clustered by zip codes. We saw from the models that there is significant unobserved heterogeneity between the *GPs* that explains why some GPs have higher ratings than others. Also, for most of the regressions we find small unobserved variation between *zip* codes where GPs have their practice. The multilevel analysis shows that a part of the unobserved variance that influences the responses is at the level of GP and zip code. The advantage of analyzing with a multilevel model is that it is more conservative and therefore provides more robust coefficient estimates, standard errors and confidence intervals compared to single level modeling. Single level models ignore the possible effect of clustering and can therefore report over-simplistic and misleading results. By including these random effects we have reduced the chance of overstatement of statistical significance (Goldstein, 2011).

4.3 Further research and policy implications

It is common for studies to report findings that indicate that people in general are very satisfied with their GP (B. Williams et al., 1998). The current study confirms this tendency. Legelisten caused great concern among the physicians to begin with; that it would serve as a public pillory of GPs (Danielsen et al., 2013). Most of the ratings are positive. The fact that these results, obtained from a private rating site with self-selected anonymous responders, have generated findings that are aligned with systematically collected data such as the Norwegian survey of living conditions, is interesting. First of all, this improves the reliability of the results regarding the association between patient satisfaction and whether they are enlisted with a GP who face patient shortage or full lists. Second, it offers confirmation of the validity of such

rating sites, whether the provided information captures what it is supposed to capture, and that the information is trustworthy. International studies have found an association between online rating sites, systematic surveys and clinical indicators (Bardach et al., 2013; Greaves et al., 2012). Their findings combined with the current results indicate that such rating sites might be useful for the public and that it is possible to trust the people, not just the experts, also when it comes to quality in healthcare (Surowiecki, 2004). This could be important for the public as the patients have imperfect information when they wish to choose a new GP because they have changed their address, or switch GP because they are unhappy with the current one. If there are conducted more studies regarding the association between patient satisfaction and clinical quality, it might contribute to involve the actual user of healthcare in planning and measuring quality to a larger extent. There should be conducted further studies on these satisfaction ratings to monitor their development over time. It should be continually compared to the national survey, as has been done in this study, and with an inclusion of more covariates that might have an influence on satisfaction. This is in order to capture whether the similar satisfaction distribution continues, as more ratings will be published. Since we only included published ratings in this study, the general high satisfaction might decrease if the dissatisfied portion in the future complies to the qualitative rules of rating, as the website and its rules for submitting are more known.

A more recent service offered by Legelisten is the ability to sign up on a list to be alerted if a specific GP has an opening on the list. It could be interesting to see whether the size of these lists is higher for the GPs with above average satisfaction ratings. It could also be interesting to see whether GPs experienced a significant change in the switching activity after the rating site was established. These objectives could tell us how people take advantage of the website and whether they consider the ratings as useful information when choosing their GP, reducing the problem of imperfect information of GPs. The rating site also provides qualitative comments for each rating. Aharony and Strasser (1993) critique the lack of prioritizing the qualitative comments that are often included in the surveys. The contents of these comments could be analyzed to see how they relate to the quantitative ratings and can provide information that can improve the ethnographic perspective and give us more details on the negative experiences (Aharony & Strasser, 1993). It could further be interesting with a panel study on the development of the ratings of the individual

GPs. For instance, if the GPs who initially have received poor ratings on certain dimensions, show improvement on those specific dimensions, this could suggest an acceptance of the ratings from the GPs point of view. It could also provide some implications on whether more published information could have the effect of increased quality, and point to what could serve as useful quality indicators.

It can be suggested that patient shortage could be such a quality indicator. Iversen and Lurås (2002) argued with their results that only GPs, who are considered as inferior and experience patient shortage, are willing to reduce their waiting times. Iversen and Lurås (2011) further commented that patients might detect inferior physician quality, which is the reason why some GPs experience patient shortage. Patients enlisted with such a GP are 50% more likely to switch and this suggests that patient shortage might predict the level of switching the GP will experience. In addition, economic theory has implied that physicians of higher quality have a higher probability of being chosen by consumers as demand is affected by the perceived quality (Biørn & Godager, 2010). Satisfaction-studies have further provided evidence that patients are less satisfied with the GPs with patient shortage, except on waiting times (Godager & Iversen, 2010, 2014; Lurås, 2007), which the results in the current study support. If we accept patient satisfaction as a valid measure for quality in healthcare (Donabedian, 1988; Pascoe, 1983; Vuori, 1991), then the satisfaction-studies might imply that lower satisfaction with GPs with patient shortage, indicates that their service is of lower quality. Therefore patient shortage could be an indicator of physician quality. However, this suggestion needs to be investigated more carefully. A GP can have patient shortage without that being associated with lower quality. For instance, the GP can prefer a larger list size than what is actually possible if located in low-populated areas. The GPs have a possibility to influence whether they are referred to as having patient shortage because they set their preferred list size themselves.

4.4 Conclusion

The results in this analysis have shown strong effects, and are aligned with and contribute to previous literature based on a randomized sample. There is an association between lower patient satisfaction and patient shortage on dimensions describing the interpersonal relationship between the patient and the GP. On the other hand, it seems that GPs with patient shortage react by offering shorter waiting times, as satisfaction is more likely to be higher on those dimensions for the same

GPs. However, since it is the interpersonal relationship that corresponds with satisfaction overall, shorter waiting times does not seem to compensate for the perceived inferior quality. This might suggest that patient shortage can provide patients with additional information about the GP than merely vacancy.

The results rely on data from a private GP rating site where responders are self-selected and anonymous. This type of rating of healthcare-providers has been unconventional in the Norwegian context and the value of such a website has been questioned. Looking at the general distribution of satisfaction, we saw that the responders have rated their GPs in the same pattern known from previous systematic surveys. This might suggest that at least in the lack of something better, rating site does not produce systematically wrong judgments of quality and could be useful in alleviating the information problem patients face when choosing a GP.

5 References

Adcock, R. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Association, 95*(03), 529-546. doi: 10.1017/S0003055401003100

Aharony, L., & Strasser, S. (1993). Patient satisfaction: what we know about and what we still need to explore. *Medical care review*.

Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit, and probit models* (Vol. 45). California: Sage Publication.

Allison, P. (1999). Logistic regression using the SAS system: theory and application. *SAS Institute, Cary, NC*, 167-168.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American economic review*, 941-973.

Baker, R. (1990). Development of a questionnaire to assess patients' satisfaction with consultations in general practice. *The British Journal of General Practice, 40*(341), 487.

Bardach, N. S., Asteria-Peñaloza, R., Boscardin, W. J., & Dudley, R. A. (2013). The relationship between commercial website ratings and traditional hospital performance measures in the USA. *BMJ quality & safety, 22*(3), 194-202.

Becker, G. S. (1991). A note on restaurant pricing and other examples of social influences on price. *Journal of Political Economy, 99*(5), 1109-1116.

Biørn, E., & Godager, G. (2010). Does quality influence choice of general practitioner? An analysis of matched doctor–patient panel data. *Economic Modelling, 27*(4), 842-853.

Bjerrum, L., & Sørensen, A. (1992). Why do patients change physicians? *Ugeskrift for laeger, 154*(50), 3587-3589.

Bornstein, B. H., Marcus, D., & Cassidy, W. (2000). Choosing a doctor: an exploratory study of factors influencing patients' choice of a primary care doctor. *Journal of evaluation in clinical practice, 6*(3), 255-262.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics, 46*(4), 1171-1178.

Carr-Hill, R. A. (1992). The measurement of patient satisfaction. *Journal of public health, 14*(3), 236-249.

Chow, A., Mayer, E. K., Darzi, A. W., & Athanasiou, T. (2009). Patient-reported outcome measures: the importance of patient satisfaction in surgery. *Surgery, 146*(3), 435-443.

Danielsen, K., Kjøllesdal, G., & Bjertnæs, Ø. (2013). *Online-vurderinger- sammenheng med systematiske brukerundersøkelser og kliniske målinger?* Oslo: Nasjonalt kunnskapssenter for helsetjenesten: Retrieved from http://www.kunnskapssenteret.no/publikasjoner/online-vurderinger-sammenheng-med-systematiske-brukerunders%C3%B8kelser-og-kliniske-m%C3%A5linger [27.03.2014].

Donabedian, A. (1966). Evaluating the quality of medical care. *The Milbank memorial fund quarterly*, 166-206.

Donabedian, A. (1988). The quality of care: How can it be assessed? *Jama, 260*(12), 1743-1748.

Dranove, D., & Satterthwaite, M. A. (1992). Monopolistic competition when price and quality are imperfectly observable. *The RAND Journal of Economics*, 518-534.

Dranove, D., & Satterthwaite, M. A. (2000). The industrial organization of health care markets. *Handbook of health economics, 1*, 1093-1139.

Dranove, D., & White, W. D. (1986). Agency and the organization of health care delivery. *Inquiry: a journal of medical care organization, provision and financing, 24*(4), 405-415.
Finnvold, J. E., Paulsen, B., & Lurås, H. (2003). De fleste fikk den de ville ha. *Samfunnsspeilet.* 5, from http://www.ssb.no/helse/artikler-og-publikasjoner/de-fleste-fikk-den-de-ville-ha [28.03.2014]
Ford, R. C., Bach, S. A., & Fottler, M. D. (1997). Methods of measuring patient satisfaction in health care organizations. *Health care management review, 22*(2), 74-89.
Gandhi, I., Parle, J., Greenfield, S., & Gould, S. (1997). A qualitative investigation into why patients change their GPs. *Family practice, 14*(1), 49-57.
Gaynor, M. (2006). Competition and quality in health care markets. *Foundation and trends in microeconomics, 2*(6), 441-508.
Gaynor, M., Moreno-Serra, R., & Propper, C. (2013). Death by Market Power:Reform, Competition, and Patient Outcomes in the National Health Services. *American Economic Journal: Economic Policy, 5* (4), 134-166.
Godager, G. (2012). Birds of a feather flock together: A study of doctor–patient matching. *Journal of health economics, 31*(1), 296-305.
Godager, G., & Iversen, T. (2010). Brukernes erfaringer med fastlege-ordningen 2001-2008-Trender i bruk, tilgjengelighet og fornøydhet *HERO* (Vol. 1): Oslo University, Health Economics Research Programme.
Godager, G., & Iversen, T. (2014). Brukernes erfraringer med fastlegeordningen 2001-2012. Trender i bruk, tilgjengelighet og fornøydhet *HERO* (Vol. 2): Oslo University, Health Economics Research Programme.
Godager, G., Iversen, T., & Ma, C. A. (2012). Competition, Gatekeeping, and Health Care Access: Oslo University, Health Economics Research Programme.
Godager, G., Iversen, T., & Ma, C. T. A. (2009). Service motives and profit incentives among physicians. *International Journal of Health Care Finance and Economics, 9*(1), 39-57.
Godager, G., & Lurås, H. (2009). Dual job holding general practitioners: the effect of patient shortage. *Health economics, 18*(10), 1133-1145.
Goldstein, H. (2011). *Multilevel statistical models*. University of Bristol, UK: John Wiley & Sons, Ltd.
Gravelle, H., & Masiero, G. (2000). Quality incentives in a regulated market with imperfect information and switching costs: capitation in general practice. *Journal of health economics, 19*(6), 1067-1088.
Gravelle, H., Propper, C., & Santos, R. (2013). Does quality affect patiens`choice of doctor? Evidence from the UK. *Working paper NO. 13/306.*
Greaves, F., Pape, U. J., King, D., Darzi, A., Majeed, A., Wachter, R. M., & Millett, C. (2012). Associations between web-based patient ratings and objective measures of hospital quality. *Archives of internal medicine, 172*(5), 435-436.
Grytten, J., Sørensen, Rune J. (2009). Patient choice and access to primary physician services in Norway. *Health Economics, Policy and Law, 4*, 11-27.
Gaardsrud, P. Ø. (2012). *Styringsdata for fastlegeordningen, 4 kvartal 2012*. Helsedirektoratet: Retrieved from http://www.helsedirektoratet.no/finn/Sider/results-finn.aspx?k=styringsdata for fastlegeordningen [01.04.2014].
Hall, J. A., & Dornan, M. C. (1990). Patient sociodemographic characteristics as predictors of satisfaction with medical care: a meta-analysis. *Social science & medicine, 30*(7), 811-818.
Hedeker, D., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 933-944.

Hekkert, K. D., Cihangir, S., Kleefstra, S. M., van den Berg, B., & Kool, R. B. (2009). Patient satisfaction revisited: a multilevel approach. *Social science & medicine, 69*(1), 68-75.
Howie, J., Porter, A., Heaney, D., & Hopton, J. (1991). Long to short consultation ratio: a proxy measure of quality of care for general practice. *The British Journal of General Practice, 41*(343), 48.
Iversen, T. (2004). The effects of a patient shortage on general practitioners' future income and list of patients. *Journal of health economics, 23*(4), 673-694.
Iversen, T., & Lurås, H. (2000). Economic motives and professional norms: the case of general medical practice. *Journal of Economic Behavior & Organization, 43*(4), 447-470.
Iversen, T., & Lurås, H. (2002). Waiting time as a competitive device: an example from general medical practice. *International Journal of Health Care Finance and Economics, 2*(3), 189-204.
Iversen, T., & Lurås, H. (2011). Patient switching in general practice. *Journal of health economics, 30*(5), 894-903.
Kann, I. C., Biørn, E., & Lurås, H. (2010). Competition in general practice: prescriptions to the elderly in a list patient system. *Journal of health economics, 29*(5), 751-764.
Le Grand, J. (2007). *The other invisible hand: Delivering Public Services through Choice and Competition*. Princeton, New Jersey: Princeton University Press.
Long, S. J. (1997). *Regression models for categorical and limited dependent variables* (Vol. 7). London: Sage Publications.
Long, S. J., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata*. Texas: StataCorp LP.
Lurås, H. (2004). *Individuals' preferences for GPs. Choice analysis from the establishment of a list patient system.* University of Oslo. (1)
Lurås, H. (2007). The association between patient shortage and patient satisfaction with general practitioners. *Scandinavian journal of primary health care, 25*(3), 133-139.
McCarthy, T. R. (1985). The competitive nature of the primary-care physician services market. *Journal of health economics, 4*(2), 93-117.
McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B, 42*(2), 109-142.
McGlone, T. A., Butler, E. S., & McGlone, V. L. (2002). Factors influencing consumers' selection of a primary care physician. *Health marketing quarterly, 19*(3), 21-37.
McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of mathematical sociology, 4*(1), 103-120.
Meltzoff, J. (1997). *Critical Thinking About Research*. Washington DC: American Psychological Association.
Midi, H., Sarkar, S. K., & Rana, S. (2010). Collinearity diagnostics of binary logistic regression model. *journal of interdisciplinary mathematics, 13*(3), 253-267.
Morrell, D., Evans, M., Morris, R., & Roland, M. (1986). The" five minute" consultation: effect of time constraint on clinical content and patient satisfaction. *British medical journal (Clinical research ed.), 292*(6524), 870-873.
Mueller, C. D. (1985). Waiting for physicians' services: Model and evidence. *The Journal of Business, 58*(2), 173-190.
Newbold, P., Carlson, W. L., & Thorne, B. M. (2013). *Statistics for Business and Economics* (Vol. 8). England: Pearson Education Limited 2013.
O`Connell, A. A. (2006). *Logistic regression models for ordinal response variables* (T. F. Liao Ed.). London: Sage publication.
Ot.prp.nr.99. (1998-99). *Om lov om endringer i lov 19. november 1982 nr.66 om helsetjenesten i kommunene og i visse andre lover (fastlegeordningen).*

http://www.regjeringen.no/nb/dep/hod/dok/regpubl/otprp/19981999/otprp-nr-99-1998-99-.html?id=160124 [01.05.2014].
Pascoe, G. C. (1983). Patient satisfaction in primary health care: a literature review and analysis. *Evaluation and program planning*, *6*(3), 185-210.
Pauly, M. V. (2004). Competition in medical services and the quality of care: concepts and history. *International Journal of Health Care Finance and Economics, 4*(2), 113-130.
Pauly, M. V., & Satterthwaite, M. A. (1979). The pricing of primary care physicians' services: a test of the role of consumer information.
Pike, C. (2010). An empirical analysis of the effects of GP competition. http://mpra.ub.uni-muenchen.de/27613/1/MPRA_paper_27613.pdf [05.11.2013]
Propper, C., Burgess, S., & Green, K. (2004). Does competition between hospitals improve the quality of care? Hopsital death rates and the NHS Internal Market. *journal of public economics, 88*, 1247- 1272.
Rabe-Hesketh, S., & Skrondal, A. (2003b). Multilevel logistic regression for polytomous data and rankings. *Psychometrica, 68*(2), 267-287.
Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modelling Using Stata*. Texas: Stata Press.
Scott, A. (2000). Economics of general practice. *Handbook of health economics*, *1*, 1175-1200.
Scott, A., & Vick, S. (1999). Patients, Doctors and Contracts: An Application of Principal ‐ Agent Theory to the Doctor ‐ Patient Relationship. *Scottish Journal of Political Economy, 46*(2), 111-134.
Sitzia, J. (1999). How valid and reliable are patient satisfaction data? An analysis of 195 studies. *International Journal for Quality in Health Care, 11*(4), 319-328.
Sitzia, J., & Wood, N. (1997). Patient satisfaction: a review of issues and concepts. *Social science & medicine, 45*(12), 1829-1843.
Sivertsen, C. T. (2014). Hvor tilfredse er pasientene med fastlegene?- en beskrivelse av anonyme brukervurderinger på en åpen internettside. *HELED skriftserie, 2*.
Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling*. London, New York, Washington, D.C: Chapman & Hall/CRC.
Sloan, F. A., & Lorant, J. H. (1977). The Role of Waiting time: Evidence from Physicians`Practices. *journal of Business*(50), 486-507.
Sullivan, M. (2003). The new subjective medicine: taking the patient's point of view on health care and health. *Social science & medicine, 56*(7), 1595-1604.
Surowiecki, J. (2004). The wisdom ofcrowds. *Why the ManyAre Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies andNations. Boston: Little, Brown*.
Train, K. (2003). Discrete choice methods with simulation. *Cambridge University Press*.
Vuori, H. (1991). Patient satisfaction—does it matter? *International Journal for Quality in Health Care, 3*(3), 183-189.
Ware, J. E., Snyder, M. K., Wright, W. R., & Davies, A. R. (1983). Defining and measuring patient satisfaction with medical care. *Evaluation and program planning, 6*(3), 247-263.
Williams, B. (1994). Patient satisfaction: a valid concept? *Social science & medicine, 38*(4), 509-516.
Williams, B., Coyle, J., & Healy, D. (1998). The meaning of patient satisfaction: an explanation of high reported levels. *Social science & medicine, 47*(9), 1351-1359.
Williams, S. J., & Calnan, M. (1991). Key determinants of consumer satisfaction with general practice. *Family practice, 8*(3), 237-242.
Winter, G. (2000). A comparative discussion of the notion of validity in qualitative and quantitative research. *The qualitative report, 4*(3), 4.

Wolinsky, F. D., & Steiber, S. R. (1982). Salient issues in choosing a new doctor. *Social science & medicine, 16*(7), 759-767.

6 Appendix

Multilevel binary logistic regression for the dependent variables; *overall, advice, listening, insight, consultation, booking* and *waiting* are presented below.

Table 6. Multilevel binary regression models for overall

	Overall	
Predictors	**exp (b)**	**95% Confidence Interval**
Patient shortage	0.613**	(0.499, 0.754)
Log herfindahl	0.994	(0.925, 1.067)
GP Female	0.828**	(0.720, 0.952)
GP age 41-50	0.764**	(0.634, 0.921)
GP age 51-60	0.592**	(0.496, 0.708)
GP age 60+	0.523**	(0.425, 0.644)
Patient Female	1.086	(0.979, 1.204)
Patient age 31-40	1.335**	(1.173, 1.521)
Patient age 41-50	1.562**	(1.357, 1.798)
Patient age51-60	1.960**	(1.656, 2.320)
Patient age 60+	2.265**	(1.866, 2.749)
County dummies±	yes	-
Constant	3.140**	(2.111, 4.672)
Random Effect	**Variance**	**Standardd error**
$\boldsymbol{\delta}_{jk}^{(2)}$ GP	1.098	0.101
$\boldsymbol{\delta}_{k}^{(3)}$ zip	0.105	0.043
Log Likelihood	-6317.19	-
NO responses	11262	-
NO GPs	3250	-
NO Zip	893	-

Significant different from zero at **1% level and *5%level.
± non are significant at a 5% level.

Table 7. Multilevel binary regression models for time use

	Booking		Waiting		Consultation	
Predictors	**exp (b)**	**95% CI**	**exp (b)**	**95% CI**	**exp (b)**	**95% CI**
Patient shortage	1.135	(0.936, 1.378)	1.405**	(1.113, 1.774)	0.796*	(0.656, 0.966)
Log herfindahl	0.932*	(0.869, 0.999)	0.996	(0.917, 1.081)	1.029	(0.966, 1.097)
GP Female	0.717**	(0.631, 0.814)	0.732**	(0.623, 0.859)	0.853*	(0.751, 0.969)
GP age 41-50	0.910	(0.771, 1.073)	0.873	(0.713, 1.069)	0.792**	(0.669, 0.939)
GP age 51-60	0.943	(0.806, 1.103)	0.690**	(0.567, 0.840)	0.627**	(0.534, 0.737)
GP age 60+	0.960	(0.797, 1.157)	0.671**	(0.531, 0.849)	0.670**	(0.554, 0.812)
Patient female	0.986	(0.890, 1.092)	0.755**	(0.669, 0.851)	1.179**	(1.070, 1.300)
Patient age 31-40	1.137	(0.996, 1.298)	0.923	(0.785, 1.084)	1.170**	(1.034, 1.324)
Patient age 41-50	1.070	(0.929, 1.233)	1.095	(0.925, 1.296)	1.458**	(1.275, 1.666)
Patient age51-60	0.884	(0.748, 1.044)	1.230*	(1.016, 1.488)	1.387**	(1.188, 1.620)
Patient age 60+	0.887	(0.733, 1.072)	1.444**	(1.169, 1.784)	1.908**	(1.592, 2.288)
County dummies±	yes	-	yes	-	yes	-
Constant	0.281**	0.191	0.161**	(0.100, 0.259)	2.870**	(1.994, 4.131)
Random Effect	**Variance**	**SE**	**Variance**	**SE**	**Variance**	**SE**
$\boldsymbol{\delta}_{jk}^{(2)}$ GP	0.4429	0.07	0.974	0.121	0.86	0.084
$\boldsymbol{\delta}_{k}^{(3)}$ Zip	0.2192	0.045	0.244	0.071	0.05	0.035
Log Likelihood	-5786.2	-	-4715.5	-	-6700	-
NO responses	10494	-	10556	-	10877	-
NO GPs	3189	-	3188	-	3217	-
NO Zips	885	-	886	-	889	-

Significant different from zero at **1% level and *5%level.
± non are significant at a 5% level.

Table 8.Multilevel binary regression models for the interpersonal relationship

	Advice		Insight		Listening	
Predictors	**exp (b)**	**95% CI**	**exp (b)**	**95% CI**	**exp (b)**	**95% CI**
Patient shortage	0.677**	(0.564, 0.813)	0.708**	(0.589, 0.851)	0.568**	(0.461, 0.700)
Log herfindahl	0.997	(0.938, 1.060)	1.023	(0.961, 1.088)	1.010	(0.940, 1.086)
GP Female	0.871*	(0.771, 0.985)	0.883*	(0.781, 0.997)	0.877	(0.761, 1.010)
GP age 41-50	0.873	(0.743, 1.026)	0.840*	(0.715, 0.987)	0.729**	(0.604, 0.881)
GP age 51-60	0.743**	(0.637, 0.867)	0.710**	(0.608, 0.828)	0.584**	(0.488, 0.700)
GP age 60+	0.734**	(0.611, 0.882)	0.688**	(0.573, 0.826)	0.520**	(0.421, 0.642)
Patient female	1.073	(0.974, 1.183)	1.035	(0.940, 1.139)	1.089	(0.980, 1.211)
Patient age 31-40	1.066	(0.943, 1.205)	1.096	(0.971, 1.237)	1.134	(0.993, 1.294)
Patient age 41-50	1.354**	(1.185, 1.547)	1.376**	(1.207, 1.569)	1.553**	(1.343, 1.795)
Patient age51-60	1.348**	(1.154, 1.574)	1.471**	(1.261, 1.715)	1.646**	(1.388, 1.953)
Patient age 60+	1.943**	(1.615, 2.338)	1.981**	(1.654, 2.373)	1.998**	(1.639, 2.436)
County dummies±	yes	-	yes	-	yes	-
Constant	2.607**	(1.841, 3.693)	2.298**	(1.622, 3.258)	4.221**	(2.806, 6.350)
Random Effect	**Variance**	**SE**	**Variance**	**SE**	**Variance**	**SE**
$\boldsymbol{\delta}_{jk}^{(2)}$ GP	0.663	0.0746	0.693	0.074	1.068	0.104
$\boldsymbol{\delta}_{k}^{(3)}$ zip	0.05	0.031	0.065	0.032	0.113	0.045
Log Likelihood	-6518.7	-	-6728. 7	-	-6052.5	-
NO responses	10904	-	10831	-	10912	-
NO GPs	3217	-	3214	-	3220	-
NO Zip	890	-	891	-	891	-

Significant different from zero at **1% level and *5%level.
± non are significant at a 5% level.